

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p mode:

Run on: November 12, 2003, 06:43:44 ; Search time 26 Seconds
(without alignments)
941.7133 Million cell updates/sec

Title: US-09-735-251-3

Perfect score: 2302
Sequence: 1 GCGAGAGGCCAGAGAGAA.....TGCCTCCATCTCCCTGTC 1273

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 566616

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn21/USPTO.spool.p/US09735251/tunat.12112003.061636.28155/app.query.fasta.1.1415
-DB=PIR_76 -OPMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPTCI=0 -LOOPT=0
-UNITS=bits START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=text -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09735251 @CGN.1.1.38 @tunat.12112003.061636.28155 -NCPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPB=OCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_76:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	189.5	8-2	373	2 S54583	ubiquitin-like pro
2	165.5	7-2	839	2 T04859	extensin homolog P
3	164.5	7-1	536	2 B84549	probable ubiquitin
4	163.5	7-1	1162	2 JH0557	exo-alpha-sialidas
5	162	7-0	815	2 B56708	extracellular sign
6	159	6-9	528	2 I47141	gastric mucin (clo
7	152	6-6	354	2 T38404	yeast dak2 homolog
8	147	6-4	1188	2 S49915	extensin-like prot
9	146.5	6-4	930	2 D37271	A-alpha 2 4 protei
10	146	6-3	351	2 S50754	hypothetical prote
11	145	6-3	3421	1 WZBBE6	367K tegument prot
12	144.5	6-3	551	2 C84549	probable ubiquitin
13	144.5	6-3	1468	2 A44345	nucleoporin - rat
14	144	6-3	1400	2 T31555	hypothetical, rat

15	142.5	6-2	1006	2 T42731	atrophin-1 related
16	142.5	6-2	3530	2 A59266	unconventional myo
17	141.5	6-1	742	2 F84643	hypothetical prote
18	141	6-1	671	2 T02504	hypothetical prote
19	140.5	6-1	1487	1 EDBEE1	immediate-early pr
20	139.5	6-1	1093	2 I38533	AF17 protein - hum
21	139.5	6-1	1844	1 RRPPTM	genome polyprotein
22	139	6-0	2715	2 T13049	eyelid - fruit fly
23	138.5	6-0	1494	2 T14355	protein-tyrosine-p
24	138	6-0	383	2 S32975	gene BCRP2 protein
25	138	6-0	872	2 S33015	hypothetical prote
26	138	6-0	992	2 A31666	hypothetical prote
27	138	6-0	1249	2 T14150	vesicle associated
28	137.5	6-0	1487	1 EDBEF6	155K transcription
29	136.5	5-9	383	2 T46707	proteophosphoglyca
30	136	5-9	1572	2 S45251	SNF2alpha protein
31	134.5	5-8	785	2 A82123	chemotaxis protein
32	134.5	5-8	1125	2 B41206	microtubule-associ
33	134	5-7	753	2 JC2099	glutelin, high mol
34	134	5-8	1647	2 S45252	SNF2beta protein -
35	133.5	5-8	419	2 T29266	hypothetical prote
36	133.5	5-8	502	2 T18562	hypothetical prote
37	133.5	5-8	574	2 T43556	Wiskott-Aldrich sy
38	133.5	5-8	574	2 T38819	Wiskott-Aldrich sy
39	133.5	5-8	907	2 E86636	hypothetical prote
40	133.5	5-8	1147	2 T42627	ADP-ribosylation f
41	133	5-8	576	2 T36729	probable serine/th
42	133	5-8	744	2 T35192	glutelin ABC trans
43	133	5-6	789	2 A30843	glutelin high mole
44	132.5	5-8	708	2 D96711	hypothetical prote
45	132.5	5-8	1952	2 T48814	hypothetical prote

ALIGNMENTS

RESULT 1

S54583
ubiquitin-like protein DSK2 - Yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YMC21.02; protein YMR276w
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 24-Sep-1999
A;Accession: S54583; S59344
R;Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54582
A;Accession: S54583
A;Molecule type: DNA
A;Residues: 1-373 <PEA>
A;Cross-references: EMBL:Z49704; NID:G825540; PIDN:CAA89774.1; PID:G825542; MIPS:YMR276
R;Biggins, S.; Rose, M.D.
submitted to the EMBL Data Library, April 1995
A;Description: The yeast ubiquitin-like proteins are involved in spindle pole body dup
A;Reference number: S59343
A;Accession: S59344
A;Molecule type: DNA
A;Residues: 1-108, 'R', 110-295, 'R', 297-373 <BIG>
A;Cross-references: EMBL:L40587; NID:G786149; PIDN:AAB07267.1; PID:G786151
C;Genetics:
A;Gene: SGD:DSK2
A;Cross-references: SGD:S0004889; MIPS:YMR276w
A;Map position: 13R
C;Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology
F;3-77/Domain: ubiquitin homology <UEH>

Alignment Scores:
Pred. No.: 1.64e-05 Length: 373
Score: 189.50 Matches: 85
Percent Similarity: 40.16% Conservative: 64
Best Local Similarity: 22.91% Mismatches: 123
Query Match: 8.23% Indels: 100
DB: 2 Gaps: 12

US-09-735-251-3 (1-1273) x S54583 (1-373):

```
QY 230 CTGATCTACTGTGTCGCGAGCTAAAGATGACGACACACCTTGACTTCTATGGATTCAA 289
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
44 LeuIleTyrSerGlyLysIleLeuLysAspGlnThrValGlySerTyrHisIleGln 63
QY 290 CCGGCTGCACCTGTCATGTTCTGCGAAAGTCTCGCTGAACTGATGACAAACCG--- 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
64 AspGlyHisSerValHisLeuVal---LysSerGlnProLysProGlnThrAlaSerAla 82
QY 346 ----- 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
83 AlaGlyAlaAsnAlaThrAlaThrGlyAlaAlaAaGlyThrGlyAlaThrProAsn 102
QY 347 ----- 347
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
103 MetSerSerGlyGlnSerAlaGlyPheAsnProLeuAlaAspLeuThrSerAlaArgTyr 122
QY 380 TTCCCGGTGTTGCACACT----- 406
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
123 AlaGlyTyrLeuAsnMetProSerAlaAspMetPheGlyProAspGlyValAlaLeuAsn 142
QY 407 AGCAGCTCTCTTACAGGAGCGGCTTTAAAG--- 439
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
143 AsnAspSerAsnAsnGlnAspGlnLeuLeuAaGlnMetMetGluAsnProIlePheGlnSer 162
QY 440 ----- 487
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 GlnMetAsnGluMetLeuSerAsnProGlnMetLeuAspPheMetIleGlnSerAsnPro 182
QY 488 GGCCTCAGCAGT---GACCTATTGCTCTGAGGTCTCCAGGACAAAGACCTCTCTCT 544
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
183 GlnLeuGlnAlaMetGlyProGlnAlaArgGlnMetLeuGlnSerProMetPheArgln 202
QY 545 GTCTTCGCTATCCCAATATGTTTGAACGTGTGGCTGCTGCTATCCAGCCCTCTGCAAT 604
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
203 MetLeuThrAsnProAspMet----- 664
QY 605 GCATTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 664
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
212 GlnSerMetGlnPheAlaArgMetMetAspProAsnAlaGlyMetGlySerIleGly 724
QY 665 TCTTCCCGGAGATGCTCCAGCTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 724
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 AlaAlaSerAlaPheProAla----- 741
QY 725 GGCCTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 784
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
242 ----- 784
QY 785 ACTCCAGCTCCGCGCCAGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 844
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
258 ThrGlyAsnAsnIleGlyThrAsnAlaGlyThrAsnAlaGlyThrAsnAlaGlyThr 844
QY 845 ACCAGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 904
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
274 ----- 904
QY 905 ACCAGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 964
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 LeuLeuAsnProAlaLeuAsnProAlaLeuAsnProAlaLeuAsnProAlaLeuAsn 964
QY 965 CAGTCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1021
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 ProAlaPheAspProAlaLeuAsnProAlaLeuAsnProAlaLeuAsnProAlaLeu 1081
QY 1022 CAGGCTCTCTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1081
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 GlnAlaGlnAspThrArgProProGlnAlaArgTyrGlnIleGlnIleLeuArgGlnLeuAsn 1340
QY 1382 GACATGGGCTATCCAGGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
141 AspMetGlyPhePheAspPheAspPheAspPheAspPheAspPheAspPheAspPheAsp 1440
```

```
QY 1141 ATCAAGCAGCCCTGGAGCTCATCTTTGCTGGA 1173
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361 ValGlnGlyAlaLeuAspSerLeuLeuAsnGly 371
```

RESULT 2

T04859
extensin homolog F28A21.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C:Accession: T04859
R:Bevan, M.; Mueller, M.W.; Muendlein, A.; Peiber, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15387
A:Accession: T04859
A:Molecule type: DNA
A:Residues: 1-839 <BEV>
A:Cross-references: EMBL:AL035526
A:Experimental source: cultivar Columbia; BAC clone F28A21
C:Genetics:
A:Map position: 4
A:Introns: 623/3
A:Note: F28A21.80

Alignment Scores:
Pred. No.: 0.000873 Length: 839
Score: 165.50 Matches: 97
Percent Similarity: 34.86% Conservative: 25
Best Local Similarity: 27.71% Mismatches: 121
Query Match: 7.19% Indels: 107
DB: 2 Gaps: 14

US-09-735-251-3 (1-1273) x T04859 (1-839)

```
QY 312 TGCAGAAAGTCTGCGCTG---AACCATGATCAGAACCGAACCTGTGCACAAAGTGGCTG 368
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 CysGlnLeuProArgLeuGluAsnPheThrPheSerTyrAsnPheThrGly----- 354
QY 369 CCATGAGAGAGTCCCGGTGTTGCACACTGCCC-----TGCA 407
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 -----GluProProValCysLeuGlyLeuProGlyPheAspArgAsnCysLeu 372
QY 408 GCACTCTCTTACAGGAGCGGTCTTTAAGATGCTCAGCAATAGAGTCTCTGATC 467
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
373 ProAlaArgProAlaGlnArgSerProGlyGlnCysAlaAlaPheSerSerLeu-ProPr 392
QY 468 AGATCATTTGCCAC-----CCAG 488
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
392 oValAspCysGlySerPheGlyCysGlyArgSerThrArgProProValValProSe 412
QY 489 CCCTCAGCAGTGCACCTTATGCTCTGCGGTCTCTCCAGGACAGGAGCTCTCTCTCT 548
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
412 rProProThrPro-----SerProGlySerPro----- 423
QY 549 TCGCTGATCCCAATATGCTTATAGTTGGTCTGCTCAGCCCTCTGTAATGCCA 608
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
424 -----ProSerProSerIleSerPr 430
QY 609 TTGCTCTGTTCTGCACTCCGTAGCAGGACTGCCCAATCCCTGGAGTCACTCTCTT 668
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
430 oSerPro-ProIle-----ThrValProSerProProThr-ThrProSer 444
QY 669 CCGGAGCATGCCCTCCAGCTCATACCGGATATGCCAGGTGGCTTCTCTTTGAAGGC 728
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 ProGlyGlySerProProSerProSerProSerIleVal----- 455
QY 729 TCTCAGATGATGAGATGACTTTCCCAAAACACAGGTCCACACCTCTAGCAGTACTC 788
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
456 -----ProSerProProSerThrThr 462
QY 789 CCAGTCCCGCCAGCTCCCTGCGGTACAGTGTGGGCGCGCGCGCGCGCGCGCGCGCG 848
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
463 ProSerProGlySerProProThrThrProThr-----ProGlyGlySerPro 481
```


RESULT 6
 147141
 Gastric mucin (clone PGM 2A) - pig (fragment)
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 03-Nov-2000
 C:Accession: 147141; S55315
 R:Turner, B.S.; Bhaskari, K.R.; Hadzopoulos-Cladaras, M.; Specian, R.D.; LaMont, J.T.
 Gastroenterology 106, 253, 1994
 A:Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a novel
 A:Reference number: 147141; MUID:94102478; PMID:7506218
 A:Accession: 147141
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-528 [147141]
 A:Cross-references: EMBL:U0281; NID:991525; PID:AAC48526.1; PID:G915208
 R:Turner, B.S.; Bhaskari, K.R.; Hadzopoulos-Cladaras, M.; Specian, R.D.; LaMont, J.T.
 Biochem. J. 308, 89-94, 1995
 A:Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.
 A:Reference number: S55315; MUID:95275264; PMID:755593
 A:Accession: S55315
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-528 [755593]
 A:Cross-references: GB:U0281; NID:991525; PID:AAC48526.1; PID:G915208
 C:Superfamily: pig submaxillary mucin

Alignment Scores:	Pred. No.	Length:
Score:	3,00236	528
Percent Similarity:	159.00	Matches: 91
Best Local Similarity:	37.02%	Conservative: 53
Query Match:	23.39%	Mismatches: 132
DB:	6.91%	Indels: 115
	2	Gaps: 15

US-09-735-251-3 (1-1273) x 147141 (1-528)

QY	378	AGTTCGCGGTGTCACACTGCCTGCACA-----GCA	470
DB	24	SerSerSerValProIleProSerThrThrSerValGlnProSerSerGlySer	43
QY	411	GCTCCTTTACAGGGGCGGCTTTTAAGATGCTGAGCAATAAGGAGTCTCTGATCAGA	470
DB	44	AlaProThrThrSerAlaThrSerValGlnThrSerSerSerSer	59
QY	471	TCATTGCGCCAGCCCGGCTCAGCAGTCAACCTATTGCTTGGGGTCTCCAGACA	530
DB	60	-----ProProIleSerSerThrIleSerValGlnThrSerSerSerSerVal	76
QY	531	AGGACCTCTCTGCTTCTGCTGATCCCAATATGCTTGATAGCTTGCTGCTCCTCACC	590
DB	77	ProThrThrSerThrThrSerValGlnPro-----Ser	87
QY	591	CAGCCCTCGTCAATGCCATTGCTCTG-----GTTCTGCATCCGTAGCAGGCGAGT	640
DB	88	SerSerSerSerAlaProThrThrArgAlaThrSerValGlnSerSerSerSer	107
QY	641	CCCCCAATGCGCTGGAGCT-----GACTCCTCTTCCCGGAGCATGCCCTCCAGC	698
DB	108	AlaProIleSerSerThrThrSerValGlnProSerSerSerGlySerValProThrThr	127
QY	699	TCATACCGGATATGCCAGGTGGCTTCTGTTGAAGGCTCTCAGATGATGAGGATGAC	748
DB	128	SerAlaThrSerValGlnSerSerSerSerAlaProThrThrSerAlaThrSer	147
QY	749	TTTACCACCAACACAGGTCCACACC-----	775
DB	148	ValGlnProSerSerSerSerSerProIleSerSerThrValSerValGlnProSer	167
QY	776	---TCTACGACACTCTCCAGCTCCCGCCAGCTCCCTCGGGTACAGTGCAGTCTCGGG	832
DB	168	SerSerSerSerAlaProThrThrSerAlaThrSerValGlnProSerSerSerSer	187
QY	833	CCC-----	835


```

Db 298 AlaSerProAlaProAlaGlnAspThrArgProProGluLeuArgTyrA:AGluGln 317
Qy 1067 CTCAGCAGCTACGTGACATGGGATCCAGAGCATAGAGCTAGCCTGGC-GCCCTGCGAG 1125
Db 318 LeuSerGlnLeuAsnGluMetGlyPheValAspPheGluArgAsnValGlnA:AlaLeuArg 337
Qy 1126 GCACCGGTGGGACATCCCAAGCAGCCTGAGACTCATCTTTGCT 1170
Db 338 ArgSerGlyGlyAsnValGlnGlyA:AlaGluSerLeuLeuSer 352

RESULT 8
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12 May-1995 #text_change 29-Oct-1999
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1108 <RUB>
A:Cross-references: EMBL:Z34465; NID:9600117; PIR:CAA84230.1; PDB:9600118

Alignment Scores:
Pred. No.: 0.0181 Length: 1188
Score: 147.00 Matches: 11
Percent Similarity: 34.33% Conservativity: 36
Best Local Similarity: 25.58% Mismatches: 170
Query Match: 6.33% Indels: 115
DB: 2 Gaps: 20

```

US-09-735-251-3 (1-1273) x S49915 (1-1188)

```

Qy 98 CCAAGTCTATTCTTCGGTGGCCAGACAGAACTGGAGAACTACTGCTAGAGGGGTAT 157
Db 262 ProLysSerIleGlyArgMetValGlyThr:LeuAspGlu 274
Qy 158 AGTATTTCTATTCTGAAGCAGCTTATGCTGGCAAACTCCAGAGAGTCTGTTCCAGAGCCT 217
Db 275 ---IleIlePheLeuAsnLysLeuAspGlyCysLeu 287
Qy 218 GAGCTGATGATCTATCTACTGCTGGTGGAAATTAAGATGACCAATACACTGACTTC 277
Db 288 LeuGluMetGlyLeu:---LeuValAsnThrThrValIleAspVal 301
Qy 278 TATGGCATTCAACTGGGTGACCTATGATGCTGTGTAATATCTGGCTGAACTGAT 337
Db 302 SerGlyAsnMetLeuValGlyThre 310
Qy 338 CAGAAACCGGAACCTGGGACAAAGTGGCTGCATGAGAGAGTTCCGGGTGTTCACACT 397
Db 311 ---ProGluGlnLeuSerAsnIleAlaLysLeuGlnLeuAspVal 325
Qy 398 GCCCTGCACAGCAGCTCTCTTACAGGAGAGCGGCTTTAAGATGCTCAACCAATAGAG 457
Db 326 ---SerArgAsnValPheThrGly:LeuValH:sgLysSerIleCysGluLeuPro 342
Qy 458 TCTCTGATCATCATTTGTGGCAGCCAGGCTCAGAGCTGACCCCTATTGCTCTTGGG 517
Db 343 AlaLeuValAsnPheSerPheAlaPheAsnPheAsnSerGlu:---AlaAlaValCys 361
Qy 518 GTTCTCCAGGACAGGACCTCTCTCTGCTCTGCTGATCCCAATATGCTGTGATACCTG 577
Db 362 MetProSerAspLysAlaLeuVal:AsnLeuAspAspArgAspAsnCysLeuGlyAlaLeu 381
Qy 578 GTGCTGCTGCAC 616
Db 382 ArgProAlaGlnLysThrAlaLeuGlnCysAlaProValLeuAlaArgProValAspCys 401

```

```

Qy 617 GTTCTGCACTCGTAGCAGGC-----AGTGCCCAATG 649
Db 402 SerLysHisValCysAlaGlyTyrProThrProGlyGlyGlyProProSerSerProVal 421
Qy 650 CCTGGAGCTGACTCTCTTCCCGAGCATGCGCTCCACTCATACCGGATATGCCAGT 709
Db 422 ProGlyLysProAlaAlaSerAlaProMetProSer----- 433
Qy 710 GCGTTCTCTGTTTGAAGAGGCTCTCAGATGATGAGGATGACTTTCACCCAAACACAGGTCC 769
Db 434 -----ProHisThrProProAspValSerProGlu-ProLeuPr 446
Qy 770 ACACCTCTAGCAGTACTCCAGCTCCCGCCAG----- 803
Db 446 GGIuProSerProVal---ProAlaProAlaProMetProMetProThrProHisSerPr 465
Qy 804 -----CCTCCCTGGGGTACAGTGGAGCTGCTGGGCCCCGGCCCATC 844
Db 465 oProAlaAspAspTyrValProProThrProProVal-----ProGlyLysSe 481
Qy 845 ACCGAGGTGAGCTGGCCACCGCTTGGCCCTGGCCGACACTCCGAGAGAGCTCTCTAC 904
Db 481 rProProAlaThrSerProSerProGlnValGlnProPro-----AlaAlaSerTh 498
Qy 905 ACACGAGCTCTGGCAGCCAGGGTCACTCTCTCAGGAGCCTCACCAGTGTCTCTG----- 959
Db 498 rPro-----ProProSerLeuValLysLeuSerProProGlnAlaProValGlySe 515
Qy 960 -----GTGTCCAGTCAGGAGCGCCATCCCAATGATCTCTTACGCAAGCC 1006
Db 515 rProProProProValLysThrSerProProAlaPro---IleGlySer-ProSerP 534
Qy 1007 CTACAGCATGCCCTTCAGGCTCTGGCAGCCAGCCAGCTTCAGAGCAGTGGGAGCCCGAG 1066
Db 534 ro-----ProProProValSerValValSerProProProValLysSerProP 551
Qy 1067 CTGCAGCAGCTACGTGACATGGGATCCAGGACGATGAGCTGAGCTCGGCCCTGCAGG 1126
Db 551 roProProAlaProValGlySerProProProProGluLysSerProProProAla 570
Qy 1127 CCA-----CCGCTGGGAGCATCCAGAGCCCTCGAGCTCATCTTTGCTGGAGGAG 1177
Db 571 ProValAlaSerProProProProValLysSerProProProThrLeuValAlaSer 590
Qy 1178 CCCATGAACCTCCCTGCTTCCCTGAGCCCGAGAGAGTTCGAGAGGCTACTGCCCTTGG 1237
Db 591 ProProProProValLysSerProProProProAlaProValAlaSer-----ProPro 608
Qy 1238 GAGGCACTCATGAAGTGGCTCATCTCTCCCTGTC 1273
Db 609 ProProValLysSerProProProProThrProVal 620

RESULT 9
D37271
A:alpha 2.4 protein - bracket fungus (Schizophyllum commune)
C:Species: Schizophyllum commune
C:Date: 07-Feb-1992 #sequence_revision 07-Feb-1992 #text_change 23-Feb-1997
C:Accession: D37271
R:Ullrich, R.C.
submitted to the Protein Sequence Database, October 1991
A:Reference number: A37271
A:Accession: D37271
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-930 <ULL>
C:Keywords: DNA binding; nucleus; transcription regulation

Alignment Scores:
Pred. No.: 0.019 Length: 930
Score: 146.50 Matches: 94
Percent Similarity: 34.82% Conservativity: 62
Best Local Similarity: 20.98% Mismatches: 154
Query Match: 6.36% Indels: 138

```



```

QY 723 AAGGGCTCTCAGATGATGAGTACTTTCCACCAACACCCAGGTCCACACCCCTCTAGCA 782
Db 188 -----ATCCAGAGTCCCGCCCA-----GCCCTCCCTGGGGTACAGTGGAGCTGCTG 830
QY 783 GT-----ACTCCAGAGTCCCGCCCA-----GCCCTCCCTGGGGTACAGTGGAGCTGCTG 830
Db 197 erProLysAlaSerProSerProSerProSerProSerProSerProSerProSerPro 217
QY 831 GGCCCCGGCCATCACCCAGAGTGGAGTGGCCCGCTGGCCCTGGCCCGTGGCCAGCTCCGG 890
Db 217 erProAlaProSerProGlnProSerProSerProSerProSerProSerProSerPro 236
QY 891 AGACCACTCTCACACACCCAGCTCCGCGACCCAGGTCATCTCTCAGGGACCTCCACCA 950
Db 236 laSerProGlnGlnSerProProSerProSerProSerProSerProSerProSerPro 254
QY 951 TGTCTCTGTGTCTCAGTCCAGGACCGCCATCACCAATGATCTCTCTCAGGCAAGCC 1007
Db 255 ThrProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 274
QY 1008 TACAGCATGCCC-----TTCAGGCTCTGGGACCCAGCTTTCAGAGCC 1052
Db 275 SerLeuSerProLysValSerProSerProSerProSerProSerProSerProSerPro 294
QY 1053 AGTGGCAGCCCGAGCTCCAGCAGCTAGCTGACATGGGCATCCAGGACGATGACTGAGCC 1112
Db 295 SerGlySerProArgAlaSer----- 301
QY 1113 TCGCGGCTCTCAGCCACCCGCTGGGAGATCCAGCAGCCCTGGAGCTCATCTTGC--- 1169
Db 302 -----ProProGly-----GlyGlyProProAlaMetSerProArgGlySerPro 316
QY 1170 -----TGGAGAGCCCGATGAGCTGAGCTGCTCTCCCTGAGC-----GCCAGCAA 1214
Db 317 ThrLeuProProThrArgSerPro-----IleLeuProProIleProSerPro 333
QY 1215 GTT---CCAGAGGCTACTGCCCTGGGAGGCA.TCATGAGGTGCCCTC 1253
Db 334 IleArgThrProSerProAlaIleGlyArgGlyArgGlyArgGlyArgGlyArgGly 349

RESULT 11
WZBB6
367K tegument protein - equine herpesvirus 1 (strain AE4p)
C:Species: equine herpesvirus 1
A:Note: host: Equus caballus (domestic horse)
C:Date: 30-Sep-1992 #sequence_revision 30 Sep 1992 #text change 16 Jul 1999
C:Accession: G36797
R:Telford, E.A.R.; Watson, M.S.; McBratney, K.; Davison, A.J.
submitted to GenBank, March 1992
A:Description: The DNA sequence of equine herpesvirus-1
A:Reference number: A36805
A:Accession: G36797
A:Molecule type: DNA
A:Residues: 1-3421 <TEL>
A:Cross-references: GB:M86664; MID:g330791; FID:AA802459.1; PID:g330816
R:Telford, E.A.R.; Watson, M.S.; McBratney, K.; Davison, A.J.
Virology 189, 304-316, 1992
A:Title: The DNA sequence of equine herpesvirus-1
A:Reference number: A41831; MJD:92295566; PMID:318606
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Genetics:
A:Gene: 24
C:Superfamily: varicella-zoster virus gene 22 protein

Alignment Scores:
Pred. No.: 0.0285 Length: 3421
Score: 145.00 Matches: 102
Percent Similarity: 33.26% Conservative: 55
Best Local Similarity: 21.61% Mismatches: 156
Query Match: 6.30% Indels: 149
DB: 1 Gaps: 22

```

```

US-09-735-251-3 (1-1273) x WZBB6 (1-3421)
QY 140 TACTCGTAGGGGCTATAGTATTTCATTCTGAGCAGCTTATTCCTGGCAAACTCCAG 199
Db 2198 TyrThrThrThrGlyThrSerThrThrPhe-----ThrLeuAlaGlySerAlaPro 2214
QY 200 GAGTCTGTTCAGAGCCCTGAGCTGATTCATCTGATCTACTCTGCT 244
Db 2215 SerAlaValLysGluValSerAlaLeuAspValAlaThrCysAlaLeuMetPheGlyAla 2234
QY 245 -----CGAAGCTAAAGATGACCAGACACTTCTGACTTC 277
Db 2235 ProValValIleAlaMetGluThrProGluMetPheSerGluAlaSerGlyMetSerPhe 2284
QY 278 -----TATGGCATTAACCTGGTCCACTGTCCATGTTCTG- 313
Db 2255 CysLeuLysIlePheAspSerArgProGlyAlaThrAspHisGluIleGlnAlaVal 2274
QY 314 -----CGAAGTCTCTGGCTGAACCT-----GATCAGAACCGGAACCTGTGGAC 358
Db 2275 SerSerAspLeuSerSerProGlyThrSerLeuLeuAlaLeuAspProAsnAlaIleGlu 2294
QY 359 AAAGTGGCTGCATGAGAGAGTCCGGGTGTTCACACTGCCCTGCACAGCAGCTCCCTCT 418
Db 2295 AsnAlaCysLeuThrThrGlnLeuGluIleLeu---SerGlyLeuValAlaSerLysLeu 2313
QY 419 TACAGGGAGGGC-----GTCTTTAAGATG 442
Db 2314 LeuAlaProAlaProCysLeuIleValLeuAspProSerMetArgValIleLysVal 2333
QY 443 CTCACANTAAGGAGTCT-----CTGGATCAGATC 472
Db 2334 LeuTrpGluSerGluSerProProAsnAspLeuValIleThrLeuAlaGluAspGluIle 2353
QY 473 ATTGTGGCCACCCAGGCTCAGCAGT----- 499
Db 2354 IleAlaGluLeuProTyThrLeuAsnAlaAspAspAspLeuLeuProProMetAsnProAsp 2373
QY 500 GACCTATGTCTTGGGGTCTCCAGGACAGGACCTC-----TTC 541
Db 2374 AspProIleTyThrArgValIleSerGlyThrAsnIleProThrAlaThrThrGluGly 2393
QY 542 TCTGTCTTGGTGTATCCCAATATGCTTCATAGTCTGGTCTGCTCACCACCCCTCTCTC 601
Db 2394 SerLeuPheAlaAsp---GlnGlnLeuGluPheLeuArgProGluSer----- 2408
QY 602 AATGCCATTGTCTCTGCTCTCTGCTCAGTCCAGCAGCTGCCCAATCCCTGGGACTGAC 661
Db 2409 AsnProProPhePheAlaSerHisAspSerSerGlnSerLeuAspValProSerSerPro 2428
QY 662 TCTCTTCCCGAGCATGCCCTCCAGCTCATACCGGGATATGCCAGGTGGCTTCTGTTT 721
Db 2429 SerSerGly-----SerAspLysTyThrGluGluAspProThrGlyValTy 2444
QY 722 GAAGGCTCTCAGATGAT----- 739
Db 2445 AspAlaProValAspAspMetSerAspMetAlaMetAsnLysAlaLysAlaTrpGlnGlu 2464
QY 740 -----GAGGATGACTTTTCCACCAACACCCAGGTCCACACCTCT 778
Db 2465 TrpLeuGluAspGlyPheAlaGluAspTyThrArgGluLeuSerAsnAlaMetPro--- 2483
QY 779 AGCAGTACTCCAGCTCCCGCCAGCTCCCTCGGGGTACAGTGGAGCTGTGGGGCCCGG 838
Db 2484 -----AlaProProLys 2487
QY 839 -----CCATCACCAGAGTCCAGTGGCCACCCGCTTGGCCCTGCGCCAGCTCCGAG 892
Db 2488 ThrThrProValValGluSerLys-----GlnLysSerAsp 2499
QY 893 AGCAGCTCTCACACACCCAGCTCTCCGACCCAGGCTCATCTCAGGACCTCCACCAATG 952
Db 2500 SerValAspArgAlaProThr-LeuProProLysAlaAlaProLeuProProSerAspAl 2519

```


Db 1008 ProProPro:~SerTyr----- 1013


```
Db      278 oHisProGlnProSerLeuHisGlyGlnGlyProGlyProHis-SerLeuGlnThrG 298
Qy      1004 GCCT-----ACGATGCGCTTCAGGCTC-----TGGCAGGCG 1039
Db      299 lyProLeuLeuGlnHisProGlyProGlnProPheGlyLeuThrProGlnSerSerG 318
Qy      1040 AGCTTCAGAGCCAGTGGCAGCCCGAGCTGACAGACTAGTACATGGGCAT--- 1092
Db      318 lngGlnGlyProLeuGlyProSerProAlaAlaA-----HisProHisSerThrI 336
Qy      1093 -----CCAGCAGCATGAGCTGAGCTCGGCGCTGCGAGGCGACCGGTGGGACATC 1144
Db      336 leGlnLeuProAla-SerGlnSerAlaLeuGlnProGlnGlnProPro-----Arg 352
Qy      1145 AAGCAGCCCTGGAGCTCATCTTTGCTGGAGAGAGCGGCAT- 1183
Db      353 GluGlnProLeuProProAlaProLeuAlaMe-ProHisIleCysProProThrThr 372
Qy      1184 GAATGCTGCTTCCTCCCTGACCC-----CCAGCAAGTTCAGAGGCTAC 1228
Db      373 ProIleProGlnLeuProAlaProGlnAlaHisLysHisProHisLysSerGlyPro 392
Qy      1229 TGCCTTGGGAGGCACTCATGAAGTGCCCTCCA 1261
Db      393 SerProPheSerMetAsnAlaAsnLeuProPr 403
```

Search completed: November 12, 2003, 06:50:12
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK nucleic - protein search, using frare, plus n2p -mode!

Run on: November 12, 2003, 06:40:09 / Search time 17 Seconds

(without alignments)

7042.94: Million cell updates/sec

Title: US-09-735-251-3

Perfect score: 2362

Sequence: 1 GCGAGAGGGCCAGAGGAGAA..... TGGCTGATCTCCCTGTC 127:

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 2.5
Ygapop 10.0, Ygapext 2.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame, n2p_model -DEV=xlp
-C=/con2.1/USP01/spec p/US09735251/runat.12.12.03.061635.28.10-app query.fasta.1.14.5
-DB=SwissProt 41 -Opt=fastan -SUFFIX=rs -MINMATCH=0 -IGORCL=0 -FOGEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 TRANS=0 -ALGO=CGI -LIST=45
-DOALIGN=200 -THR SCORES=0 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MCDS=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=20000000
-USER=US09735251 -ACGN 1.16 -@runat.12.12.03.061635.28.10 -NCPJ 6 -ICPU=1
-NO.MMAP -LARGEOJURY -NEG SCORES=0 -WAIT -DSBLOCK=0 -LING=0
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELCP=6 -DELEX=7

Database : SwissProt 41.1

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the best hit being printed,
and is derived by analysis of the data table distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	188.5	8.2	373	1 DSK2 YEAST	P48510 saccharomyc
2	163.5	7.1	1162	1 TCNA TRYCR	P23253 trypanosoma
3	162	7.0	815	1 MK07 HUMAN	Q13164 homo sapien
4	159	6.9	802	1 ENAH MOUSE	Q31373 mus musculu
5	152	6.6	354	1 YANG SCHRC	Q12169 schistosach
6	148	6.4	1083	1 T2D3 HUMAN	C0C268 homo sapien
7	145	6.3	806	1 MK37 MOUSE	Q9WVS8 mus musculu
8	145	6.3	3421	1 TEGU HSVEB	P28955 equine herp
9	144.5	6.3	1468	1 N153 RAT	P49791 rattus norv
10	143.5	6.2	555	1 DAB1 HUMAN	Q75553 homo sapien
11	143	6.1	830	1 SREC HUMAN	Q14162 homo sapien
12	142.5	6.2	3530	1 MY15 HUMAN	Q9UK77 homo sapien
13	142	6.2	902	1 NFCA HUMAN	Q14534 homo sapien
14	141.5	6.1	555	1 DAB1 MACFA	Q9BGX5 macaca fasc
15	141.5	6.1	940	1 MAZ4 SCHCO	P37938 schizophyll
16	140.5	6.1	1487	1 ICP4 HSVEK	P17473 equine herp
17	139.5	6.1	1093	1 AF17 HUMAN	P55138 homo sapien
18	139.5	6.1	1844	1 POLR TYMV	P20123 turnip yell

19	138	6.0	992	1 EBN6 EBV	P03204 epstein-bar
20	137.5	6.0	1487	1 IC94 HSVEB	P28925 equine herp
21	134.5	5.8	559	1 WAS1 HUMAN	Q92558 homo sapien
22	134.5	5.8	1125	1 MAP4 MOUSE	P27546 mus musculu
23	134	5.8	1647	1 SN24 HUMAN	P51532 homo sapien
24	133.5	5.8	907	1 A180 HUMAN	O60641 homo sapien
25	133.5	5.8	1509	1 GSR1 HUMAN	Q9NZM4 homo sapien
26	133	5.8	1124	1 IRS2 HUMAN	Q9Y4H2 homo sapien
27	132	5.7	464	1 S3A2 HUMAN	Q15428 homo sapien
28	132	5.7	1198	1 HCN4 RAT	Q9JKA7 rattus norv
29	132	5.7	1586	1 SN22 HUMAN	P51531 homo sapien
30	132	5.7	1844	1 POLR TYMV	P28477 turnip yell
31	131.5	5.7	1367	1 AMYH YEAST	P08640 saccharomyc
32	131	5.7	616	1 PLR1 RABIT	P17487 oryctolagus
33	131	5.7	863	1 M113 HUMAN	O8N3F8 homo sapien
34	131	5.7	1175	1 HCN4 RABIT	Q9TV66 oryctolagus
35	131	5.7	1567	1 FMX2 MOUSE	Q9J104 mus musculu
36	131	5.7	1970	1 RPB1 HUMAN	P24328 homo sapien
37	131	5.7	1970	1 RPB1 MOUSE	P08775 mus musculu
38	130.5	5.7	743	1 TFE3 HUMAN	P19532 homo sapien
39	130.5	5.7	2161	1 SHK1 HUMAN	Q9Y566 saccharomyc
40	130	5.6	1273	1 WE91 YEAST	P38968 saccharomyc
41	130	5.6	1844	1 POLR TYMV	P10358 turnip yell
42	129.5	5.6	475	1 S3A2 MOUSE	O62203 mus musculu
43	129.5	5.6	3164	1 TEGU HSVEB	P10220 herpes simp
44	129	5.6	522	1 NG62 HUMAN	P37198 homo sapien
45	129	5.6	598	1 NR41 CANFA	P51666 canis famil

ALIGNMENTS

RESULT 1

DSK2 YEAST	ID	DSK2 YEAST	STANDARD:	PRT:	373 AA.
AC	P48510:				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DE	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	Ubiquitin-like protein DSK2.				
GN	DSK2 OR SHE4 OR YMR276W OR YW8021.02.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAN=S288C;				
RX	MEDLINE=96281973; PubMed=8682868;				
RA	Riggins S., Ivanovska I., Rose M.D.;				
RT	"Yeast ubiquitin-like genes are involved in duplication of the				
RT	microtubule organizing center."				
RL	J. Cell Biol. 133:1331-1346(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAN=S288C / AB972;				
RX	PubMed=9169872;				
RA	Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,				
RA	Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,				
RA	Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,				
RA	Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;				
RT	"the nucleotide sequence of Saccharomyces cerevisiae chromosome				
RT	XIII."				
RL	Nature 387:90-93(1997).				
CC	- FUNCTION: INVOLVED, WITH RAD23 IN SPINDLE POLE BODY DUPLICATION.				
CC	- SUBCELLULAR LOCATION: Nuclear (Probable).				
CC	- SIMILARITY: TO S.POMBE SPAC26A3.16.				
CC	- SIMILARITY: Contains 1 ubiquitin-like domain.				
CC	- SIMILARITY: Contains 1 UBA domain.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				

```

modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-s.b.ch/announce/
or send an email to license@sb-s.b.ch).
-----
EMBL: L40587; AAC07267.1;
EMBL: Z49704; CA89774.1;
PIR: S54583; S54583;
HSP: Q15843; NCD;
SGD: S0004889; DSK2;
InterPro: IPR006636; STIL;
InterPro: IPR000449; UBA_domain;
InterPro: IPR006626; Ubiquitin;
Pfam: PF00627; UBA; 1
Pfam: PF00240; ubiquitin; 1
PRINTS: PR00348; Ubiquitin; 1
SMART: SM00727; STIL; 2
SMART: SM00165; UBA; 1
SMART: SMC0213; UBQ; 1
PROSITE: PS00030; UBA; 1
PROSITE: PS00259; Ubiquitin; 1
PROSITE: PS00053; Ubiquitin; 2
KW Nuclear protein.
FT DOMAIN 1 76 Ubiquitin-LIKE
FT DOMAIN 127 171 UBA
FT CONFLICT 109 109 E -> A IN REF 21
FT CONFLICT 296 296 P -> A IN REF 21
SQ SEQUENCE 373 AA; 39516 MW; 25E0E2B39F40D56 CRC64;
-----
Alignment Scores:
Pred: No.: C:000107 Length: 373
Score: 188.50 Matches: 95
Percent Similarity: 40.16% Conservative: 44
Best Local Similarity: 22.91% Mismatches: 123
Query Match: 8.19% Indels: 100
DE: Gaps: 22
-----
US 09-735-251-3 (1:1273) x DSK2_YEAST (1:373)
CY 230 CTGATCTACTGTGTCGGAAGCTAAAGATGACACGACGACCTGACTGCTCATGGCATGAA 289
|||||
44 LeuIleTyrSerGlyLysIleLeuLysAspAspThrValGluSerTyrHisIleGln 63
CY 290 CCTGGGTCCTACTGCCAGTGTCTCCGAAAGTCTAGCTGAACTGATGATGAGAAACCG... 346
|||||
64 AspGlyHisSerValHisLeuValLysSerGlnProGlySerProGlnThrAlaSerAla 52
CY 346 .. 346
DB 83 AlaGlyAlaAsnAsnAlaThrAlaThrAlaThrAlaThrAlaThrAlaThrAlaThrAla 302
CY 347 .. 347
103 MetSerSerGlyGlnSerArgGlyPheAsnProLeuAlaAspLeuThrSerAlaAspTyr 122
CY 380 TTCGGGTGTGCACACT..... ..GCCCTGAC 406
|||||
123 AlaGlyTyrLeuAsnMetProSerAlaAspMetPheGlyProAspGlyGlyAlaSerAsn 142
CY 407 AGCAGCTCCTCTTACAGGGAGGCCGTCTTTAAG..... 439
|||
143 AsnAspSerAsnAsnGlnAspGluLeuLeuArgMetMetGluAsnProIlePheGlnSer 162
440 .. 440
-----ATGCTCAGCAATAAGAGATGTCTGATCAGATCATTTGGCCACCCCA 487
|||||
163 GlnMetAsnGluMetLeuSerAsnProGlnValLeuAspPheMetIleGlnSerAsnPro 182
CY 488 GGCCTCAGCACT--GACCCCTATTGCCCTTGGGTTCTCCAGGACCAAGGACCTCTTCTCT 544
|||
183 GlnLeuGlnAlaMetGlyProGlnAlaArgGlnMetLeuGlnSerProMetPheArgGln 202
CY 545 GTCTTCGCTGATCCCAATATCTCTGTAACGTTGGTGGCTGCTCACCAGGCCCTGTCGAAT 604
|||
203 MetLeuThrAsnProAspMet.....IleArg 211
|||

```

CC alpha-(2-8)-glycosidic linkages of terminal sialic residues in
 CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
 CC synthetic substrates.
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POSSIBLE).
 CC -!- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOXASTIGOTES, MINIMUM
 CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
 CC -!- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
 CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA
 CC PROTEIN.
 CC -!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
 CC -!- SIMILARITY: Contains 3 BNR repeats.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M61732; AAA130255.1; -
 CC PIR: JH0557; JH0557
 CC InterPro: IPR02860; GH_BNR
 CC Pfam: PF02012; BNR; 2
 CC Hydrolyase; Glycosidase; Glycoprotein; Repeat; GPI anchor;
 CC Phosphorylation.
 CC KW DOMAIN 1 457 CYS-RICH.
 CC FT REPEAT 23 34 BNR 1
 CC FT REPEAT 163 174 BNR 2
 CC FT REPEAT 209 220 BNR 3
 CC FT DOMAIN 458 588 FIBRONECTIN TYPE-III.
 CC FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS, LTP DOMAIN.
 CC FT CARBOHYD 342 342 N-LINKED (GLCNAC) (POTENTIAL).
 CC FT CARBOHYD 394 394 N-LINKED (GLCNAC) (POTENTIAL).
 CC FT CARBOHYD 1125 1125 N-LINKED (GLCNAC) (POTENTIAL).
 CC SQ SEQUENCE 1162 AA; 120032 MW; 07C49221B97C6A42 CRC64;
 CC
 CC Alignment Scores:
 CC Pred. No.: 0.00469 Length: 1162
 CC Score: 163.50 Matches 109
 CC Percent Similarity: 36.16% Conservative: 49
 CC Best Local Similarity: 24.94% Mismatches: 181
 CC Query Match: 7.10% Indels: 98
 CC DB: 1 Gaps: 19
 CC
 CC US-09-735-251-3 (-1273) x TNCA_TRYC -11162
 CC QY 59 TGCACCTGCGGTGAGCTGCTTAAAGCTTA TTTA TTTAATTTTCGGTGG 118
 CC Db 499 TyrHisLeuValLeuThrMetAlaAsn..... 507
 CC QY 119 CCAGACAGAACTGGGAGAACTACTCGTAGGCGGATATAGTATTTCATTCTGAGAGCAG 178
 CC Db 508 -----LysileGlyserValTyrIleAspGly 517
 CC QY 179 CTTATTCTGGCAACTCCAGAGAGTCTGTCCAGAGAGCTTAAATATTTGAT 229
 CC Db 518 LeuLeuGluGlySerGlyGlnThrValValProAspGlyArgThrProAspIleSerHis 537
 CC QY 229 ----- 229
 CC Db 538 PheTyrValGlyGlyTyrLysArgSerAspMetProThrIleSerHisValThrValAsn 557
 CC QY 230 -----CTGATCTACTGTGGTGGAGCTAAAGATGACAGACACTTGACITCTATGCG 283
 CC Db 558 AsnValLeuLeuTyrAsnArgArgGlnLeuAsnThrGlnGlnLeuArgThrLeuPheLeu 577
 CC QY 284 ATTCAACCT-----GGGTCCACTGTCCCTATGTCTCGAAGAGCTCGGCTGAACTGAT 317
 CC Db 578 SerGlnAspLeuIleGlyThrGluAlaHisMetAspSerSerAspSerAlaHis 597
 CC QY 339 CAGAAACCGGAA-----CCTGTGGACAAAGTGGCTGGCATGAGAGAGTTCGGTGTGCAC 394

Db 598 SerThrProSerThrProAlaAspSerSerAlaHisSerThrProSerThrProValAsp 617
 QY 395 ACTGCTGTGACAGCAGCTCTCTTTACAGGAGCGGTCTTTAAGATGCTCAGCAATAAG 454
 Db 618 SerSerAlaHisSerThrProSer----- 625
 QY 455 GAGTCTCTGGATCAGATCATGTGGCCACCCAGCCCTCAGCAGCTGACCTTATGCTCTT 514
 Db 626 ThrProAlaAspSerSerAlaHisGlyThrProSerThrProValAspSerSerAlaHis 645
 QY 515 GGGTCTCTCCAGGACAGGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 571
 Db 646 GlyThrProSerThrProAlaAspSerSerAlaHisGlyThrProSerThrProValAsp 665
 QY 572 ACGTTGGTGGCTGTCTCCACCCAGCCCTCTCTCAATGCCATTGCTCTCTCTCTCTCTCT 631
 Db 666 SerSerAlaHisSerThrProSer-----ThrProValAspSerSer 679
 QY 632 GCAGGAGTGGCCCAATGCCCTGGGACTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 685
 Db 680 AlaHisSerThrProSerThrProValAspSerSerAlaHisGlyAlaProSerThrPro 699
 QY 686 -----ACCTCATACCCGGATATGCCAGTGGCTCTCTCTCTCTCTCTCTCTCTCTCT 739
 Db 700 AlaAspSerSerAlaHisGlyThrPro-----SerThrPro 711
 QY 740 GAGATGACTTTTCCACCCAAACACCCAGGTCCACACCTCTAGCAGTACTCTCCAGCTCCCGC 799
 Db 712 ValAspSerSerAlaHisGlyThrProSerThrProAlaAspSerSerAlaHisSerThr 731
 QY 800 CCAGCTCTCTCTGGGTACAGTGGAGCTGCTGGGCGCCGCCATCCACCCAGCAGTGGCTG 859
 Db 732 ProSerThrProAlaAspSer---SerAlaHisSerThrProSerThrProAlaAsp--- 749
 QY 860 GCCACCGCTTGGCCTGGCCAGCAGCTCCGAGAGAGCAGCTCTCACACCCAGCTCTCTGGC 919
 Db 750 SerSerAlaHisSerThrProSerThrProValAspSerSerAlaHisGlyThrProSer 769
 QY 920 ACCAGGCTGATCTCTCAGGACCTCCACCAATGCTCTCTCTCTCTCTCTCTCTCTCTCT 970
 Db 770 ThrProAlaAspSerSerAlaHisSerThrProSerThrProAlaAspSerSerAlaHis 789
 QY 971 GGGAGCGCCATCA---CCAATGATCTCTTTCAGCAAGCCCTACAGCATGCCCTTTCAGGCC 1027
 Db 790 GlyThrProSerThrProValAspSerSerAla-----HisSerThrProSerThrPr 807
 QY 1028 TCTGGCGCCAGCAGCTTCAGAGCCAGTGGCAGCCAGCTCCAGCCAGCTGCGAGCTACGTG 1087
 Db 807 oValAspSerSerAlaHisGlyThrProSerThrProValAspSerSer----- 823
 QY 1088 GGCATCCAGGACGATGAGCTGAGCTGCGGCGCTTCAGGCCAGCCAGCCAGCTGGGACATCCA 1147
 Db 824 AlaHisSerThrProSerThrProValAspSerSerAlaHis-----GlyThrProSe 841
 QY 1148 CAGCCCT-----GGAGCTCATCTTTGCTGGAGAGCCCATGAATCCCTCTCTCTCTCC 1198
 Db 841 rThrProValAspSerSerAlaHisSerThrProSerThrProAlaAspSerSerAlaHi 861
 QY 1199 CCTGAACCCCGCAGCAAGTTGGAGAGGCTACTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1247
 Db 861 sSerThrProSerThrProAlaAspSerSerAlaHisGlyThrProSer 877
 CC
 CC RESULT 3
 CC MK07 HUMAN
 CC ID MK07 HUMAN STANDARD; PRT: 815 AA.
 CC AC Q13164; Q16634;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Mitogen-activated protein kinase 7 (EC 2.7.1.1.) (Extracellular signal-
 CC DE regulated kinase 5) (ERK-5) (ERK4) (BMK1 kinase).
 CC GN MAPK7 OR PRK7 OR ERK5 OR ERK4.

OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

RN 1.
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=9527943; PubMed=7759517;
 RA Zhou G, Bao Z-Q, Dixon J E.
 RT "Components of a new human protein kinase signal transduction
 pathway.";
 RL J. Biol. Chem. 270:12665-12669 (1995);

RN 2.
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95374539; PubMed=7646528;
 RA Lee J-D, Ulevitch R-J, Han J.
 RT "Primary structure of BMK1: a new mammalian map kinase.";
 RL Biochem. Biophys. Res. Commun. 213:715-724 (1995).

CC FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND
 CC NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.
 CC ENZYME REGULATION: Activated by tyrosine and threonine
 CC phosphorylation (by similarity).
 CC TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN
 CC HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE
 CC IN LIVER.
 CC DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN
 CC TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.
 CC PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES. WHEN
 CC THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY
 CC ROLE, IS ABSENT.
 CC SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MAP KINASE SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@sib-sib.ch).

CC EMBL: U25278; AAA81381.1;
 CC EMBL: U29725; AAA82931.1;
 CC EMBL: U29726; AAA82932.1;
 CC EMBL: U29727; AAA82933.1;
 CC PIR: B56708; R56708.
 CC KSSP: P24941; IRLC.
 CC Genew: HGNC:16980; MAPK7.
 CC MIM: 602521;
 CC GO: GO:0004707; F:MAP kinase activity; TAS
 CC GO: GO:0007165; P:signal transduction; TAS
 CC InterPro: IPR003527; MAP_kin.
 CC InterPro: IPR000719; Prot_kinase.
 CC InterPro: IPR002290; Ser_thr_kinase.
 CC Pfam: PF00069; pkinase; 1.
 CC ProDom: PD000001; Prot_kinase; 1.
 CC SMART: SMG0220; S_TKc; 1.
 CC PROSITE: PS01351; MAPK; 1.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00109; PROTEIN_KINASE_DOM; 1.
 CC PROSITE: PS00109; PROTEIN_KINASE_ST; 1.
 CC Transferrase; Serine/threonine-protein kinase. ATP binding. Cell cycle;
 CC Phosphorylation.
 CC FT DOMAIN 54 346 PROTEIN KINASE.
 CC FT DOMAIN 337 340 POLY-ALA.
 CC FT DOMAIN 433 464 PRO-RICH 1.
 CC FT DOMAIN 520 523 POLY-ARG.
 CC FT DOMAIN 577 699 PRO-RICH 2.
 CC FT NP_BIND 60 68 ATP (BY SIMILARITY).
 CC FT BINDING 83 83 ATP (BY SIMILARITY).
 CC FT ACT_SITE 181 181 BY SIMILARITY.
 CC MOD_RES 218 218 PHOSPHORYLATION (ACTIVATES THE KINASE)

FT MOD_RES 220 220 (BY SIMILARITY).
 FT PHOSPHORYLATION (ACTIVATES THE KINASE)
 FT (BY SIMILARITY).
 FT AREGTRPHRCLCS -> GPVVEPANTAASVA (IN
 FT REF. 2).
 FT CONFLICT 609 609 L -> V (IN REF. 2).
 SQ SEQUENCE 815 AA; 88636 MW; 379AD69803207CCF CRC64;
 Alignment Scores:
 Pred. No.: 0.00555 Length: 815
 Score: 182.00 Matches: 89
 Percent Similarity: 35.69% Conservative: 27
 Best Local Similarity: 27.38% Mismatches: 118
 Query Match: 7.04% Indels: 91
 DB: Gaps: 13
 US-09-735-251-3 (1-1273) x MKC7_HUMAN (1-815)
 QY 311 CTCGAAAGTCTCGCTCAACCTGATCAGAACCGGAACCTGTGCAAAAGTGGCTGCC 370
 DB 490 LeuArgAspGlyProSerAlaProLeuGluAlaProGluProArgLysProValThrAla 509
 QY 371 -----ATGAGACAGTTCGGGTGTTGCACACTGCCCTGCACAGCAGCTCTCTTAC 421
 DB 510 GlnGluArgGlnArgGluArgGluGluLysArgArgGlnGluArgAlaLysGlu 529
 QY 422 AGGAGGCGGCTTTAAGATGCTCAGCAATAAGGAGTCTCTCGATCAGATCATCTTGGGCC 481
 DB 530 ArgGlu-----LysArgArgGlnGluArgGluArgGluArgGluArgGluAla 546
 QY 482 ACCCGAGGCTCAGCAGTACCTATTGCTCTTGGGGTTCTCCAGGACAAGACCTTCTTC 541
 DB 547 SerGlyGlyProSerThrAspProLeuAlaGlyLeuValLeuSerAspAsn----- 563
 QY 542 TCTGTCTTCGCTGATCCCAATATGCTTGATACGTTGGTG-----CCTGCTCAC 589
 DB 564 -----AspArgSerLeuLeuGluArgTrpThrArgMetAlaArgProAlaAla 579
 QY 590 CCAGCCCTGTCATCAATGCCATTGCTCTGCTTGTGCTACCTCCGTAGCAGCAGTGCCTCAATG 649
 DB 580 ProAlaLeuThr-----SerValProAlaProAlaProAla 591
 QY 650 CTGGGAGTCACTCTCTT-----CCGAGAGATCCCTCCAGCTCATACCGGAT 700
 DB 592 ProThrPro ThrProThrProValGlnProThrSerProProGly----- 607
 QY 701 ATGCCAGTGGCTCTCTGTTGAAGGGCTCTCAGATGATGAGGATGCTTTCACCCCAAC 760
 DB 608 -----ProLeuAlaGlnPr 612
 QY 761 ACCAGTCCACACCTCTTAGCAGTA-----CTCCAGCTCCCGCCCA 802
 DB 612 oThrGlyProGlnProGlnSerAlaGlySerThrSerGlyProValProGlnProAlaCy 632
 QY 803 GCCTCCCTGGGTACAGTGAGCTGTGGCGCCGCCCATCACCCAGCAGTGGCTGCC 862
 DB 632 sProPro-----ProGlyProAlaProHisProThrGlyPr 644
 QY 863 ACCG---CCTTGGCTGGCCAGACCTCCGAGAGCAGCTCTCACACACCGACTCTCTGCC 919
 DB 644 oProGlyProLeuProValProAlaProProGlnIleAlaThrSerThrSerLeuLeuAl 664
 QY 920 ACCCAGGTCATTCTCAGGGAGCTCACCAGTCCCAATGTCTCTGTGTCTCCAGTCCAGGAGCC 979
 DB 664 aAlaGln-----SerLeuValProProGlyLeuPr 675
 QY 980 ATACCAATGATCTCTTACCCCAAGCCCTACAGCATGCCCTCAGGCCTCTGGGAGCC 1039
 DB 675 o-----GlySerSerThrProGlyValLeuProTrpPheProProGlyLeuPr 691
 QY 1040 AGCTCTCAGAGCCAGTGGAGCCCGCAGCTGCACAGCTAGTGCATGCGCATGGCATCCAGGAC 1099
 DB 691 oProProAspAlaGlyGlyAlaProGlnSerSer--MetSerGlu----- 705

QY 278 TATGGCAATCAACCTGGTCCACTGTCCATGTCTTCGAAAGCTCCTGGCTGAACCTGAT 337
 Ds 308 -----ProProThrSer 311
 QY 338 CAGAAACCGAAGCTGTGGACAAAGTGGCTGCATGAGAGAGTTCGGGGTGTTCACACT 397
 Ds 312 ThrProProThrProProLeuArgHisA:aaIaThrArgPheAlaThrSerLeuGlySer 331
 QY 398 GCCTGCAGCAGCAGCTCTCTTACAGGGAGG:GGTTCAGAGCTCTCAGCAATAGAGAG 457
 Ds 332 AlaPheHisProValLeuProHisA:ThrValProArgProLeu:AsnLysAsn 350
 QY 458 TCTCTGGATCAGATCATTTGGCCACCCCGAGCCTCAGCAGTGCACCTATTGCTCTGGG 517
 Ds 351 SerArgProSerSerProValAsnThrPro:SerSerGlnProProAlaAlaGly 368
 QY 518 GTTCTCCAGCAGCAGCAGCT 577
 Ds 368 sSerCysAlaThrProThrSerAsnPheSerProLeu:----- 380
 QY 578 GTGCTGCTCAGCCAGCCTCTCTCAATGCCATTGTCTGCTCTCTCTCTCTCTCTCTCTCT 637
 Ds 381 -----ProProSerProProLeuValSerSerProProGlyLysAla 395
 QY 638 A:-----GTGCTCAATGCTCTGGAGCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 679
 Ds 395 aThrGlyProArgProValLeuProValCysValSer:SerProValPro:GlnMe 413
 QY 680 CCTCTCAGCTCATACCGGATATGCCAGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 739
 Ds 413 tProProSerProThrAla:-----ProAsnGlySerLeuAs 425
 QY 740 GAGGATGATTTTACCCCAACAGCAGGTCTCAACCTCTAGCAGTACTCTCTCTCTCTCTCT 799
 Ds 425 pSerValThrThrProValSerProProProThrSerGlyProAlaA:apProProPr 445
 QY 800 CAGGCT 859
 Ds 445 oProProPro:-----ProProProProProProProProProProProProProPro 457
 QY 860 GCACCCGCTTGGCTGGCGAGCACTCTGGAGCAGCAGCTCTCTCTCTCTCTCTCTCTCTCT 904
 Ds 457 uProProPro:ProLeuProProProLeu:AlaSerLeuSerHisCysGlySerGlnAla 475
 QY 905 ACACCCGCTCTGGCAGCCAGGCT 955
 Ds 476 SerProProProGlyThrProLeuAlaSerHisCysGlySerProValLeuPro 495
 QY 956 -----TCTGCT 997
 Ds 496 SerProSerAlaGlyAlaProAlaSerAlaGlyHisProLeuAsnProLeuGlyAsp 515
 QY 998 AGCCAAAGCCTCAGCATGCT 1057
 Ds 516 SerSerAlaSerGluProGlyLeuGlnAlaAlaSerHisProAla:GluSerPro 533
 QY 1058 CAGCCGAGCTCAGCAGCT 1117
 Ds 534 ThrProGln:-----GlyLeuValLeuGly 541
 QY 1118 CCTCTGAGG:-----CCACCTGTGGGAGCATCTCAAGCAGCTCTCTCTCTCTCTCTCTCTCT 1171
 Ds 542 ProProAlaProProProProProProProProProProProProProProProProProPro 561
 QY 1172 GAGCAGCCCATCAACT 1231
 Ds 562 ProProProProProProPro:ProProProProProProProProProProProProProPro 579
 QY 1232 CCTTGGAGGCACTCATGAAGGCT 1293
 Ds 580 Pro:-----ProProProProProProProProProProProProProProProProPro 589

RESULT 5

YAU8 SCHPO STANDARD; PRT; 354 AA.
 ID YAU8 SCHPO
 AC Q10159:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein C26A3.16 in chromosome I.
 GN SPAC26A3.16.
 OS Schizosaccharomyces pombe (Fission Yeast)
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Spours J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares K., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voicikert G., Aert R., Robben J., Grymonprez B.,
 RA Weijens I., Vansteenkiste E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehtach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Jesery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 CC -!- SIMILARITY: TO YEAST DSK2.
 CC -!- SIMILARITY: Contains 1 ubiquitin-like domain.
 CC -!- SIMILARITY: Contains 1 UBA domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-aib.ch/announce/>
 CC or send an email to license@isb-aib.ch).
 CC -----
 DR EMBL; Z69240; CAA93239.1;
 DR PIR; T38404; T38404.
 DR GeneDB Sponbe; SPAC26A3.16;
 DR InterPro; IPR006636; STIL.
 DR InterPro; IPR000449; UBA domain.
 DR InterPro; IPR000626; Ubiquitin.
 DR Pfam; PF00627; UBA; 1.
 DR Pfam; PF00240; ubiquitin; 1.
 DR SMART; SM00727; STIL; 2.
 DR SMART; SM00165; UBA; 1.
 DR SMART; SM00213; UBO; 1.
 DR PROSITE; PS00303; UBA; 1.
 DR PROSITE; PS00053; UBIQUITIN_2; 1.
 KW Hypothetical protein.
 FT DOMAIN 1 78 UBIQUITIN-LIKE
 FT DOMAIN 309 353 UBA.
 FT SEQUENCE 354 AA; 36819 MW; 1A99B2D97E73A831 CRC64;

[3] SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
MEDLINE:97098442; PubMed:8942982;
Tanese N., Saiju D., Vassallo M.F., Chen J.-L., Adron A.,
"Molecular cloning and analysis of two subunits of the human TFIIID
complex: hTFII130 and hTFII100.";
Proc. Natl. Acad. Sci. U.S.A. 93(1361): 14616(1996).
CC - FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT
CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
CC BY THE AP-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
CC SUBUNIT: TFIIID IS COMPOSED OF DATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
CC SUBCELLULAR LOCATION: Nuclear.
CC SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.sib.ch/announce/
CC or send an email to licenses@sib.ch).

EMBL; Y11354; CAAT2189.1;
EMBL; AL137377; CAC36006.1;
EMBL; AL109911; CAC22312.2;
EMBL; U75308; AAC50901.1;
PCB; IH30; 26-SEP-02;
TRANSFAC; TC2328;
GeneW; HGNC:11537; TAF4.
MIM; 601796;
GO; GO:0005669; Transcription factor TFIID complex, TAS
GO; GO:0016251; General RNA polymerase II transcription fac.; TAS.
GO; GO:0005515; Protein binding activity, TAS
GO; GO:0003713; Transcription co-activator activity, TAS
InterPro; IPR003894; TAF_hom.
Pfam; PF05236; TAF4_1.
SMART; SMC0549; TAFH; 1.
Transcription regulation; Nuclear protein; D-structure.
DOMAIN 39 42 POLY-HIS
FT DOMAIN 52 57 POLY-ALA
FT DOMAIN 99 101 POLY-GLY
FT DOMAIN 142 148 POLY-ALA
FT DOMAIN 269 275 POLY-PHE
FT DOMAIN 311 337 POLY-ALA
FT DOMAIN 680 683 POLY-SER
FT DOMAIN 808 813 POLY-ALA
FT DOMAIN 828 831 POLY-ASP
FT CONFLICT 105 117 PGSPSPPELVIA > GENCODEVORGRES
(IN REF. 1)
FT CONFLICT 136 136 A->S (IN REF. 2)
FT CONFLICT 185 185 G->GGG (IN REF. 2)
FT CONFLICT 233 264 MISSING (IN REF. 3)
FT CONFLICT 293 293 P->L (IN REF. 3)
SQ SEQUENCE 1083 AA; 109943 MW; A6453E24572A275; CRC64;

Alignment Scores:
Pred. No.: 0.048 Length: 1093
Score: 148.00 Matches: 97
Percent Similarity: 31.21% Conservative: 31
Best Local Similarity: 26.30% Mismatches: 174
Query Match: 6.43% Indels: 104
DB: 1 Gaps: 15

US-09-735-251-3 (-1273) x TD3_HUMAN (-1283)

QY 355 GGACAAGTGCGTCGCCATGAGAGAGTCCGGGTCTTSCACTGCCTTCACAGCACTC 414
||| |||
Db 74 GlyProAlaProAlaGluGiYaAPrOGiValoAlaProIleProIroAlaudly 93
415 CTCTTACAGGAGCGCGTCTTTTAGTGGTGTUAGCAATAAGGATCTCTTACAGCATAT 474
||| ||| ||| |||

[illegible]

Db 2589 ProThrLeuProProAlaProProLeuProGlnSerThrSer-----Lys 2603

Qy 1241 GCACCTCATGAGGTGGCTCCATCTCTCCCT 1270

EL 2604 AlaAlaSerGlyProProThrLeuPro 2613

RESULT 9

N153_RAT STANDARD; PRT: 1468 AA.

AC P49791;

DT 01-OCT-1996 (Rel. 34, Created;

DT 01-OCT-1996 (Rel. 34, Last sequence update);

DT 28-FEB-2003 (Rel. 41, Last annotation update;

DE Nuclear pore complex protein Nup153 (Nucleoporin Np153) (153 kDa

DE nucleoporin).

GN NUP153.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN 1.

RP SEQUENCE FROM N.A., AND SEQUENCE OF 602-613, 622-645 AND 971-993.

RC STRAIN=Buffalo; TISSUE=Liver;

RX MEDLINE=91137325; PubMed=8422679;

RA Sukegawa J., Blobel G.;

R1 "A nuclear pore complex protein that contains zinc finger motifs,

R1 binds DNA, and faces the nucleoplasm.";

RL Cel: 72:29-38(1993).

CC -!- FUNCTION: Possible DNA-binding subunit of the nuclear pore

CC complex (NPC). The repeat-containing domain may be involved in

CC anchoring components of the pore complex to the pore membrane.

CC -!- SUBUNIT: Interacts with SENP2 [By similarity].

CC -!- SUBCELLULAR LOCATION: Nuclear pore complex. Located to the

CC terminal ring structure of the nucleoplasmic cage.

CC -!- DOMAIN: Contains F-X-F-G repeats.

CC -!- PTM: Phosphorylated in interphase, hyperphosphorylated during

CC mitosis. May play a role in the reversible disassembly of the

CC nuclear pore complex during mitosis.

CC -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS

CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST

CC NUP1, NSP1, POM 12, AND MAMMALIAN P62.

CC -!- SIMILARITY: Contains 4 RANBP2-type zinc fingers.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC at the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed, changed, copied and for commercial

CC entities requires a license agreement (See: http://www.ebi.ac.uk/submit/submit.html)

CC Or send an email to license@ebi.ac.uk.

CC

DR EXBL: L06821; NOT_ANNOTATED_CDS.

DR PIR: A44345; A44345.

DR InterPro: IPR001876; Znf_RangDP.

DR Pfam: PF01093; Nucleoporin_FG, 9.

DR Pfam: PF00641; zf-RanBP, 4.

DR SMART: SM00547; Znf_RBZ, 4.

DR PROSITE: PS01358; ZF_RANBP2_1, 4.

DR PROSITE: PS01093; ZF_RANBP2_2, 4.

KW Nuclear protein; Transport; Repeat; Zinc-finger; DNA-binding;

KW Phosphorylation.

FT DOMAIN 4 15 GLY-RICH.

FT DOMAIN 442 446 POLY-GLY.

FT ZN_FING 657 687 RANBP2-TYPE 1.

FT ZN_FING 721 750 RANBP2-TYPE 2.

FT ZN_FING 789 818 RANBP2-TYPE 3.

FT ZN_FING 846 875 RANBP2-TYPE 4.

SQ SEQUENCE 1468 AA; 152823 MW; C3DFD9697C556A7C CRC64;

Alignment Scores:

Pred. No.: 0.0761 Length: 1466

Score: 144.50 Matches: 106

Percent Similarity: 36.82% Conservative: 49

Best Local Similarity: 25.18% Mismatches: 144

Query Match: 6.28% Indels: 122

DB: 1 Gaps: 20

US-09-735-251-3 (1-1273) x N153_RAT (1-1468)

Qy 149 GGGGCTAT-----AGTATTTCATTCTGAGCAGCTATT 184

Db 1073 GlyGlyPheThrPheGlyLysValAspSerAlaAlaLeuSerSerProSerMetPheVal 1092

Qy 185 GCTGGCAAACTCCAGGAGTCTGTCAGACCCCTGAGCTGATTGATCTGATCTGCTGTGT 244

Db 1093 LeuGlyArgThrGluGluLysGlnGluProValThrSerThrSerLeuValPheGly 1112

Qy 245 CGGAAGCTA-----AAAGATGAC 262

Db 1113 LysLysAlaAspAsnGluGluProLysCysGlnProValPheSerPheGlyAsnSerGlu 1132

Qy 263 CAGACACTTGATCTTATGGCATTCAACTGGGTCACATGTCATGTCGGAAGTCC 322

Db 1133 GlnThrLysAspGluSerSerLysProThrPheSerPheSerValAlaLysProSer 1152

Qy 323 TGGCTCGAACCTGATCAGAAACCGGAA----- 349

Db 1153 ValLysGluSerAspGlnLeuAlaLysAlaThrPheAlaPheGlyAsnGlnThrAsnThr 1172

Qy 350 CTTGTGGCAAAAGTGGCTGCCATGAGAGAGTTCCTGGGTTTCACACTGCCCTGCACAGC 409

Db 1173 ThrThrAspGlnGlyAlaLysProAlaPheSerPheLeuAsnSerSerSerSer 1192

Qy 410 AGCTTC-----TCCTACAGGAGGCGGTCTTAAGATGCTCAGCAATTAAGGAG 457

Db 1193 SerSerThrProAlaThrSerSerSerAlaSerIlePheGlySer---SerThrSerSer 1211

Qy 458 TCTCTGGATCAGATCATTTGTGGCCACCCAGCCCTCAGAGTGACCCCTATTCTCTCTGG 517

Db 1212 SerSerProProValAlaAlaPheValPheGlyGlnAlaSerAsnProValSer----- 1229

Qy 518 GTTCTCCAGGACAAGGACCTCTCTCTCTCTGCTGATCCCAATATGCTTCATACGTTG 577

Db 1230 -----SerSerAlaPheGlyAsnSerAlaGluSerSerThrSer 1242

Qy 578 GTGCT-----GCTCACCAGCCCTCGTCAATGCCATTCCTCTGTT 619

Db 1243 GlnProLeuLeuPheProGlnAspGlyLysProAlaThrThrSer----- 1257

Qy 620 CTGCATCTGGTAGGAGGAGTCCCA-----AGCCTGGGACACTGCTCTCTCTCCCG 673

Db 1258 -----SerThrAlaSerAlaAlaProProPheValPheGlyThrGlyAlaSerSerAsn 1275

Qy 674 AGCATGCTCTCCAGCTCATACCGGATATGCCAGGTGGCTCTCTGTTTGAAGGCTCTCA 733

Db 1276 SerThrValSerSer-----GlyPheThrPheGlyAlaThrThr 1288

Qy 734 GATGATGAGGATGACTTTCACCCAAACACAGGTCACACCCCTCTAGC----- 781

Db 1289 -----ThrSerSerSerSerGlySerPheValPhe 1299

Qy 782 -----AGTACTCCAGCTCCCGCCAGCC-----TCC 808

Db 1300 GlyThrGlyHisSerAlaProSerAlaSerPheGlyAlaAsnGlnThrProThr 1319

Qy 809 CTGGGCTACAGT---GGAGCTCTGGGGCCCGCCCATCACCCAGAGTGAGTGCCACCC 865

Db 1320 PheGlyGlnSerGlnGlyAlaSerGlnProAsnProProSerPheGlySerIleSerSer 1339

Qy 866 GCCTTGGCCCTGGCCAGC-----ACTCCG 889

Db 1340 SerThrAlaLeuPheSerAlaGlySerGlnProValProProThrPheGlyThrVal 1359

Qy 890 GAGACAGCTCTCACACCCGACT-----CCTGGACCCAGGCTCATTCCTCA 937

Db 1360 SerSerSerGlnProProValPheGlyGlnGlnProSerGlnSerAlaPheGlySer 1379

QY 938 GGGACCTCACCAATGCTCTGCTCCAGTCAGGAGCCGACCATCACCAATGATCTCTC 997
 DB 1380 GlyThrAlaAsnAlaSerValPheGlnPheGlySerThrThrAsn-----Phe 1397
 QY 998 AGCCAGCCCTACAGCATGCCCTTCAGGCTC-----TGSCAGCCGAGGCTTCAGAGC 1051
 DB 1398 AsnPheThrAsnAsnAsn--ProSerGlyValPheThrPheGlyAlaSerProSe: 1415
 QY 1052 CAGTGGCAGCCGACCTGACAGCATGCTAGTGCATCGGCATCTCAGGACGATGAGCTGAGC 1111
 DB 1416 -----ThrProAlaAlaAlaGlnProSerGlySerGlyGlyPheSerPheSer 1432
 QY 1112 CTGGCGCTGACGAGCCGCTGGGACATCTT AAGCAGCCCTGAGGCTCATCT 1165
 DB 1433 GlnSerProAlaSerPheThrValGlySerAsnGlyAsnMetPheSerSerSer: 1491
 RESULT 10
 DAB1_HUMAN STANDARD; PRT; 555 AA.
 AC 075553; Q9NVAB.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Disabled homolog 1.
 GN DAB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99009343; PubMed=9790777;
 RA Lambert de Rouvroit C., Goffinet A.M.
 RT "Cloning of human DAB1 and mapping to chromosome 1p32";
 RL Genomics 53:246-247(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Fazili Z., Sun W., Xu X.-X.;
 RT "Aberrant disabled-1 expression in tumors";
 RL Submitted (MAY-2000) to the EMBL/GenBank/CDR databases.
 CC -!- FUNCTION: Adapter molecule functioning in neural development. (By similarity).
 CC -!- SUBUNIT: Associates with the SH2 domains of Src, Fyn and Abl (by similarity).
 CC -!- DOMAIN: THE PID DOMAIN STRUCTURALLY RESEMBLES THE ASN PRO-XAA TYR(P) MOTIF FOUND IN MANY TYROSINE PHOSPHORYLATED PROTEINS.
 CC -!- PTM: Phosphorylated on Tyr 148 and 149 and 150 upon cell induction in embryonic neurons (by similarity). Also phosphorylated on Ser-491 independently of cell signaling (by similarity).
 CC -!- SIMILARITY: Contains 1 PID domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)
 CC
 CC EMBL; AF071062; AAC70068.1;
 CC EMBL; AF263547; AAF73058.1;
 CC Genbank; HGNC:2661; DAB1.
 CC
 CC MIX; 603448;
 CC InterPro; IPR006020; PTB_P1D.
 CC Pfam; PF00640; PID; 1.
 CC SMART; SM00462; PTB; 1.
 CC PROSITE; PS01179; PID; 1.
 CC Developmental protein; Neurogenesis; Phosphorylation.
 CC DOMAIN 36 189 PID.
 CC MOD_RES 198 198 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 220 220 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 491 491 PHOSPHORYLATION (BY CDK5) (BY

FT CONFLICT 86 86
 FT CONFLICT 240 241
 FT CONFLICT 291 291
 FT CONFLICT 298 298
 FT CONFLICT 326 326
 FT CONFLICT 335 335
 FT CONFLICT 379 379
 FT CONFLICT 553 553
 SC SEQUENCE 555 AA; 59979 MW; A4B133A0C3AF2034 CRC64;
 Alignment Scores:
 Pred. No.: 0.077 Length: 555
 Score: 143.50 Matches: 87
 Percent Similarity: 35.94% Conservative: 51
 Best Local Similarity: 22.66% Mismatches: 143
 Query Match: 6.23% Indels: 103
 DB: 19
 US-09-735-251-3 (1 1273) x DAB1_HUMAN (1-555)
 QY 242 GGTGGGAGGATAAAGATGACGACGACATCTGATCTTATGGCATTCACTGGGTCCACT 301
 DB 57 GlyAspLeuGlySerMetMetLysLeuGlyValValAlaGlyAla--- 75
 QY 302 GTCCATCTTCGCAAGTCTCGCTGACCTGATCAGAAACCGGAACTGTG----- 355
 DB 76 -----ArgSerLysGlyGluHisGlnLysLeuPheLeuThrLeuSer 90
 QY 356 -----GACAAAGTGGCTGACGAGAGATTCGCGGTGTTG 391
 DB 91 PheGlyGlyLeuLysPheAspGluLysThrGlyAlaLeuGlnHis----- 106
 QY 392 CACACTGCGCTGACGACGACCTCTTACAGGAGGCGGCTC-----TTT 436
 DB 107 HisAlaValHisGluLeuSerTyrLeuAlaLysAspLeuThrAspHisAlaPhe 126
 QY 437 AAGATGTCAGCAATAAGGAGTCTCTGATCATCATCTGTCGCCACCCAGCCCTCAGC 496
 DB 127 GlyTyrValCysGlyLysGluGlyAsnHisArgPheValAlaLeuLysThrAlaGlnAla 146
 QY 497 AGTGACCTATTGCTCTTGGGGTCTCAGGACAGGACCTCTTCTGCTTGGCTGAT 556
 DB 147 AlaGluProVal-----IleLeuAspLeuArgAspLeuPheGlnLeuLysTyrGlu 163
 QY 557 CCAATATGTTGATACGTTG-----GTGCTGCTCAGCCAGCC 595
 DB 164 LeuLysHisArgGluGluLeuGluLysAlaGlnLysAspLysGlnCysGluGlnAla 183
 QY 596 CTCCTCAATGCCATTGTC-----CTGGTCTG 622
 DB 184 ValTyrGlnThrLeuGluGluAspValGluAspProValTyrGlnTyrIleValPhe 203
 QY 623 CACTCCCTAGCAGGACGATGCCCAATG-----CTGGGACTGACTCTCTTCCCGGAGC 676
 DB 204 Glu---AlaGlyHisGluProIleArgAspProGluThrGluGluAsnLeuTyrGln 221
 QY 677 ATGCGCTCCAGCTATACCGGGATATCCAGTGGCTTCTGTTTGAAGGCTCTCAGAT 736
 DB 222 ValProThrSerGlnLysLys-----GluGlyValTyrAsp 233
 QY 737 GATCAGGATGACTTTCACCCCAACACCGATCCACACCTCTAGCAGTACTCCAGGCTCC 796
 DB 234 -----ValProLysSer 237
 QY 797 GCGCCAGCTCTCCCTGGGGTACAGTGGAGCTGCTGGGCCCCCGGCCCATCACCCAGAGTGA 856
 DB 238 GlnProValSer-----AlaValThrGlnLeuGlu 247
 QY 857 CTGCGCCAGCGCTTGGCGCTGGCCAGCAGCTCCGAGGAGGAGCTCTCACACAGGCTCTCT 916
 DB 248 LeuPheGlyAspMet-----SerThrProAspLeuThrSerProThrPro 264

QY 917 GGCACCCAGGT-----CATTCCTCAGGACCTCACCACATGTCTCTGTGT 961
 Db 265 AlaThrProGlyAspAlaPheIleProSerSerGlnThrLeuProAaAaSerAlaAaP 284
 QY 962 GTCCAGTCAGGAGCCCATCACCACATGATCTCTCAGGCAAGCCCTACAGATGCCCT 1021
 Db 285 ValPheSerValPro-----LeuGlyThrAlaAaValProGlyTyr 300
 QY 1022 CAGCCCTCTGGCAG-----CCAGCCCTTCAAGCCAGTGGCAGCCCGAGCTGCACAG 1075
 Db 301 ValAlaMetGlyAaValLeuProSerPhePheGlyGln---GlnProLeuValGlnGln 319
 QY 1076 CTAGCTGACATGCGCATCCAGSAGATGAGCTGAGCTGGCGCTCAGGCCACCGCTG 1135
 Db 320 GlnMetValMetGlyAaHisProProValAlaGlnValMetProGlyAaGlnProIle 339
 QY 1136 GGCACATCAAGCAGCCCTGAGCTCATCTTCTGGAGGAGCCCATGAACCTCCCTGT 1195
 Db 340 AlaIrrpGly---GlnProGlyLeuPheProAlaThrGlnGlnProThrValAla 358
 QY 1196 TCCCTGTAACCCCGCAGCAAGTGGAGAGGCAATGAGCTGGAGGAGCACTCATGAAGTG 1255
 Db 359 GlyGlnPheProProAaAa-----PheValProThrGlnThrValMetProLeu 375
 QY 1256 CCTCATCTCTC 1267
 Db 376 ProAlaAaMet 379

RESULT 11
 SREC_HUMAN
 ID_SREC_HUMAN STANDARD: FRT: 830 AA
 AC Q14162: Q43370:1
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Endothelial cells scavenger receptor precursor (Acetyl LDL receptor
 DE (Scavenger receptor class F member 1).
 GN SCARF1 OR SREC OR KIAA149.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11: Genes: 272:31217-31220(1997).
 RP J. Biol. Chem. 272:31217-31220(1997).
 RN [2].
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22086180; PubMed=11978792.
 RA Adachi H., Tsujimoto M.;
 RT "Characterization of the human gene encoding the scavenger receptor
 RT expressed by endothelial cell and its regulation by a novel
 RT transcription factor, endothelial zinc finger protein-2.";
 RL J. Biol. Chem. 277:24014-24021(2002).
 RN [3].
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=96127530; PubMed=8590280;
 RA Nagase T., Seki N., Tanaka A., Ishikawa K., Nishimura N.;
 RT "Prediction of the coding sequences of unidentified human genes. IV.
 RT The coding sequences of 40 new genes (KIAA121-KIAA160) deduced by
 RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 2:167-174(1995).
 CC - FUNCTION: Mediates the binding and degradation of acetylated low
 CC density lipoprotein (Ac-LDL). Mediates heterophilic interactions.
 CC suggesting a function as adhesion protein (By similarity).
 CC - SUSUNIT: Heterophilic interaction with SREC2 via its extracellular
 CC domain. The heterophilic interaction is suppressed by the presence
 CC of ligand such as Ac-LDL (By similarity).

CC - SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC - TISSUE SPECIFICITY: Endothelial cells.
 CC - SIMILARITY: Contains 6 EGF-like domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D86864; BA24070.1;
 CC EMBL: AB052946; BAC2692.1;
 CC EMBL: D63483; BAC9770.1;
 CC HSP: P01180; 2B2.
 CC Genes: HGNC:16820; SCARF1.
 CC GO: GO:0016021; C: integral to membrane; IDA.
 CC GO: GO:0030169; F: low-density lipoprotein binding activity; IDA.
 CC GO: GO:0004888; F: transmembrane receptor activity; TAS.
 CC GO: GO:0045192; P: low-density lipoprotein catabolism; TAS.
 CC GO: GO:0006899; P: receptor mediated endocytosis; TAS.
 CC InterPro: IPR006209; EGF-like.
 CC InterPro: IPR006210; IEGF.
 CC InterPro: IPR02049; Laminin_EGF.
 CC PRINTS: PR0001; EGF_LAMININ.
 CC SMART: SM00181; EGF_5.
 CC PROSITE: PS00622; EGF_1; 6.
 CC PROSITE: PS01186; EGF_2; 6.
 CC Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
 CC EGF-like domain; Glycoprotein; POTENTIAL.
 CC SIGNAL 1 19
 CC CHAIN 20 830
 CC DOMAIN 20 421
 CC TRANSMEM 422 442
 CC DOMAIN 443 830
 CC DOMAIN 53 87
 CC DOMAIN 95 130
 CC DOMAIN 155 191
 CC DOMAIN 215 249
 CC DOMAIN 302 339
 CC DOMAIN 351 382
 CC DOMAIN 476 520
 CC DOMAIN 622 798
 CC DOMAIN 431 448
 CC DISULFID 63 75
 CC DISULFID 77 86
 CC DISULFID 99 111
 CC DISULFID 105 118
 CC DISULFID 120 129
 CC DISULFID 159 172
 CC DISULFID 165 179
 CC DISULFID 181 190
 CC DISULFID 219 230
 CC DISULFID 225 237
 CC DISULFID 239 248
 CC DISULFID 306 319
 CC DISULFID 323 326
 CC DISULFID 329 338
 CC DISULFID 355 363
 CC DISULFID 358 370
 CC DISULFID 372 381
 CC CARBOHYD 382 389
 CC CARBOHYD 389 392
 CC CARBOHYD 393 393
 CC CONFLICT 662 662
 CC R -> W (IN REF. 3).
 CC SEQUENCE 830 AA; 87430 MW; F560D9E1A64D779 CRC64;
 Alignment Scores:
 Pred. No.: 0.0875 Length: 830
 Score: 143.00 Matches: 128
 Percent Similarity: 26.79% Conservative: 40

NCBI_TaxID=9606;

SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
MEDLINE=20021762; PubMed=10532926;
Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,
Barber T.D., Miller W., Touchman J.W., Jin L., Sullivan S.L.,
Sellers J.R., Camper S.A., Lloyd R.V., Kachar B., Friedman I.B.,
Fridell R.A.,
Characterization of the human and mouse unconventional myosin XV
genes responsible for hereditary deafness DFNB3 and shaker 2.,
Genomics 61:243-258(1999).
(2)
PARTIAL SEQUENCE FROM N.A., AND VARIANTS (FN3) TTS 2111 AND PHE-2113.
MEDLINE=98267311; PubMed=9603736;
Wang A., Liang Y., Fridell R.A., Probst F.J., Wilcox E.R.,
Touchman J.W., Morton C.C., McNeil R.J., Kober-Trauth K., Camper S.A.,
Friedman I.B.,
Association of unconventional myosin MYO15 mutations with human
nonsyndromic deafness DFNB3.,
Science 280:1447-1451(1998).
-!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS BY
SIMILARITY. MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF
THE ACTIN-RICH STRUCTURES OF THE INNER EAR SENSORY HAIR CELLS.
-!- SUBCELLULAR LOCATION: Cytoplasmic.
-!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PITUITARY. ALSO EXPRESSED
AT LOWER LEVELS IN ADULT BRAIN, KIDNEY, LIVER, LUNG, PANCREAS,
PLACENTA AND SKELETAL MUSCLE. NOT EXPRESSED IN BRAIN IN THE
PITUITARY, HIGHLY EXPRESSED IN ANTERIOR GLAND CELLS.
-!- DISEASE: Defects in MYO15A are the cause of an autosomal recessive
form of nonsyndromic deafness (DFNB3).
-!- SIMILARITY: Contains 1 myosin like globular head domain.
-!- SIMILARITY: Contains 1 FERM domain.
-!- SIMILARITY: Contains 3 IQ domains.
-!- SIMILARITY: Contains 1 MYTH4 domain.
-!- SIMILARITY: Contains 1 SH3 domain.
This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation.
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed, usage by and for commercial
entities requires a license agreement. See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch.
EMBL: AF144094; AAF05931;
EMBL: AF051974; ... NOT ANNOTATED FOR
PIR: A59266; A59266.
HSP: P10587; IBR2.
Genew: HGNC:7594; MYO15A.
MIM: 602666;
MIM: 600376;
GO: GO:0007605; P:hearing; TAS.
InterPro: IPR000299; Band 4.1.
InterPro: IPR000048; IQ region.
InterPro: IPR001609; myosin_head.
InterPro: IPR000857; MYTH4.
InterPro: IPR001452; SH3.
Pfam: PF00612; IQ; 3.
Pfam: PF00663; myosin_head; 1.
Pfam: PF00784; MYTH4; 2.
PRINTS: PR00193; MYOSINHEAVY.
ProDom: PD000355; myosin_head; 1.
SMART: SM00295; B41; 1.
SMART: SM00015; IQ; 3.
SMART: SM00242; MYSC; 1.
SMART: SM00139; MYTH4; 2.
SMART: SM00326; SH3; 1.
PROSITE: PS00660; FERM_1; FALSE_NEG.
PROSITE: PS00661; FERM_2; FALSE_NEG.
PROSITE: PS00057; FERM_3; 1.

DR PROSITE: PS50096; IQ; 3.
DR PROSITE: PS500C2; SH3; FALSE_NEG.
KW Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW Calmodulin-binding; Disease mutation; Deafness.
FT DOMAIN 1888 2329 HEAD OR MOTOR DOMAIN.
FT DOMAIN 2033 3530 NECK OR REGULATORY DOMAIN.
FT DOMAIN 1888 2329 TAIL.
FT DOMAIN 2033 3530 COILED COIL (POTENTIAL).
FT DOMAIN 1792 1799 ACTIN-BINDING (POTENTIAL).
FT DOMAIN 1902 1924 IQ 1.
FT DOMAIN 1925 1954 IQ 2.
FT DOMAIN 1955 1976 IQ 3.
FT DOMAIN 2867 2953 SH3.
FT DOMAIN 3269 3530 FERM.
FT NP BIND 1315 1322 ATP (POTENTIAL).
FT VARIANT 2111 2111 N -> Y (IN DFNB3; FAMILY FROM BENGKALA).
FT VARIANT 2113 2113 /FTID=VAR_010303.
FT VARIANT 2113 2113 I -> F (IN DFNB3; INDIAN FAMILY).
FT /FTID=VAR_010304.
SQ SEQUENCE 3530 AA; 395171 MW; 3D103923D4BCB4A CRC64;
Alignment Scores:
Pred. No.: 0.115 Length: 3530
Score: 142.50 Matches: 95
Percent Similarity: 36.31% Conservative: 35
Best Local Similarity: 26.54% Mismatches: 138
Query Match: 6.19% Indels: 90
DB: 1 Gaps: 19
US-09-735-251-3 (1-1273) x MYO15_HUMAN (1-3530)
QY 296 TCACGTGTCATGTTCTGCGAAAGTCCTGGCTGAACCTGATCAGAAACCGAACCTGTG 355
Db 571 SerLeuAlaArgPheLeuLysLysThrLeuSerGlu-----LysLysProIle 586
QY 356 GACAAAGTGGTGGTCCATGACAGAGTTCGGGTGTTCACACTGCCCTGCACGACGTCC 415
Db 587 AlaArgLeuArgGlySerGlyLysThrArgAla-----GlyGlyPro 600
QY 416 TCTTACAGGAGCGGTCTTTAAGATGCTCAGCAATAAGAGTCTCTGCATCAGATCAT 475
Db 601 AlaValArgGluAlaAlaArgLysArgPheGlyTyrLys-----LeuAla 615
QY 476 GTGGCCACCCAGGCGCTCAGCAGTGCACCTATGCTCTTGGGTCTCCAGCAGCAGCAG 535
Db 616 GlyMetAspProGlyLysProGlyThrProIleValLeuArgAlaGln-ProArgAl 635
QY 536 CTTCTCTTCTGCTTGGCTG-----ATCCCAATATGCTTAT 571
Db 635 aArgSerSerAsnAspAlaArgArgProProAlaProGlnProAlaProArgThrLeu 655
QY 572 AGCTTGGTGGCTGCTCAGCCAGCCCTCGTCAATGCAATGCTGCTGCTGCTGCTGCTG 631
Db 655 rHisTrpSerAlaLeuLeuSerProProValProProArg----- 668
QY 632 GCAGCAGTGCCTCCCAATGCTGGGACTGCTCTCTTCCCGAGCAGTCCCTCCAGCTCA 691
Db 669 -----ProProSerSerGlyProProAlaPro-----ProLeuSerProAlaLe 684
QY 692 TACCGGGATATGCCAGTGGCTTCTCTGTTTGGGGCTCTCAGATGATGAGATGACTTT 751
Db 684 uSerGlyLeuProArgProAlaSerProTyr---GlySer-----LeuArgAr 699
QY 752 CACCCAAACACAGCGGTGCACACCTCTAGCAGTACTCCAGCTCCCGCCAGCTCCCTG 811
Db 699 gHisProProProTTPAlaAlaProAlaHisValProProAlaProGlnAlaSerGly 719
QY 812 GGTACAGTGGAGCTGCTGGGCGCCCGCCGCTCAACCCAGAGTGGAGCTGGCCACCG- 866
Db 719 pAlaPheAlaGlu-----ProProAlaValSerProGluVal-----ProProAspLe 735
QY 867 -----CCTTGGCTGCGCAGCAGTCTCGGAGCAGCTCTCACACACCGACTCTCT 916

QY 700 T-----ATGCCAGGTGGCTTCCTGTTTGAAGGGCTCTACAGATGATGAGGATGACTTCA 753
 Db 157 ualaglyalaglnlyg.yglyAlaPhe-----PheSe 168
 QY 754 CCCAAACACACAGGTCCACACCTCTACAGT-----ACTCCAGCTCCCGCC 801
 Db 168 rProSerProGlySerSerLeuSerSerLeuSerSerLeuSerSerPheSerPheSerAspAl 188
 QY 802 AGCTCCCTGGGTACAGT----- 830
 Db 198 uAlaAlaLeu---TyrAlaAlaCysAspGluValGluSerGluLeuSerGluAlaAla 207
 QY 821 -----GGAGCTGCTGGGCGCCGCGCATCAACACAGAGTACG-----TGGCCAC 864
 Db 207 rArgPheGlyLeuGlySerProLeu-ProSerProArgAlaSerProArgProTrepThP 227
 QY 865 CGCCTTGGCCTGCCAGCACTCCGGAGACAGAGTCTCACACACCGACTCTCGGCCACCA 924
 Db 227 roGluAspProTrp-----SerLeuTy-G.yProS 237
 QY 925 GGGTCATTCCTCAGGAGCTCACCAGTCTCTGTGTCTCAGTCCAGGACGCCCATCAC 984
 Db 237 erProGlyGlyArgGlyProGluAspSerProLeuLeuSerAlaProGlyProThP 257
 QY 985 CAATGATCTCTTACGCCAAGCCCTACAGATGCGCTTCAGAGCTCTGGGAGCCAGCCT 1044
 Db 257 ro-----AlaSerProArgProAlaSer-----ProCysGlyLysArgArg 271
 QY 1045 TCAGAGCAGTGGCGCCGCGCTGCAGCAG-----CTACGTGACATGGCATCCAGG 1097
 Db 271 yrSerSerSerGlyThrProSerSerAlaSerProAlaLeuSerArgGlySerLeuG 291
 QY 1098 ACGATGAGCTCAGCTGGCGGCTTSCAGAGTACAGGTGGGACATCCAGCAGCCCTGSA 1157
 Db 291 lyGluGlyGlySerGluProProProProProProLeuProLeuAlaArgAspProGlyS 311
 QY 1158 GC----- 1159
 Db 311 erProGlyProPheAspTyValGlyAlaProProAlaGluSerIleProGlnTyThrA 331
 QY 1160 --TATCTTCTCGAGGAGCCCATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1217
 Db 331 rGArgThrSerSerGluGlnAlaValAlaLeuProArgSer GluGluProAlaSerC 350
 QY 1218 GCAGAGGCTACTGCTGGGAGGACATCAAAATGCTCTCA 1261
 Db 350 ysAsnGlyLysLeuProGluGlyLysArgGlnProValAlaSer 364

RESULT 14
 DAB1_MACFA
 ID DAB1_MACFA STANDARD: PRT, 555 AA
 AC Q9BGX5;
 DT 28-FEB-2003 (Rel. 41, Created;
 DT 28-FEB-2003 (Rel. 41, Last sequence update;
 DT 15-SEP-2003 (Rel. 42, Last annotation update;
 DE Disabled homolog 1.
 GN DAB1;
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Frontal cortex;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hira M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Adapter molecule functioning in neural development (By
 similarity).

CC -!- SUBUNIT: Associates with the SH2 domains of Src, Pyn and Abl (By
 similarity).
 CC -!- DOMAIN: THE PID DOMAIN SPECIFICALLY BINDS TO THE ASN-PRO-XAA-
 TYR(P) MOTIF FOUND IN MANY TYROSINE-PHOSPHORYLATED PROTEINS.
 CC -!- PTM: Phosphorylated on Tyr-198 and Tyr-220 upon reelin induction
 in embryonic neurons (By similarity). Also phosphorylated on Ser-
 491 independently of reelin signaling (By similarity).
 CC -!- SIMILARITY: Contains 1 PID domain.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 or send an email to license@ebi.ac.uk).

CC EMBL: AB055282; BAB21906.1;
 DR InterPro: IPR006320; PTB_PID.
 DR Pfam: PF00840; PID.1.
 DR SMART: SM00462; PTB.1.
 DR PROSITE: PS0179; PID.1.
 KW Developmental protein; Neurogenes; Phosphorylation.
 FT DOMAIN 36..189 PID.
 FT MOD_RES 198..198 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 220..220 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 491..491 PHOSPHORYLATION (BY CDKS) (BY
 SIMILARITY).
 SQ SEQUENCE 555 AA; 59920 MW; A38002A1D0C4EE51 CRC64;

Alignment Scores:
 Pred. No.: 0.103 Length: 555
 Score: 141.50 Matches: 83
 Percent Similarity: 35.53% Conservative: 52
 Best Local Similarity: 21.84% Mismatches: 150
 Query Match: 6.15% Indels: 95
 DB: 1 Gaps: 18

US-09-735-251-3 (1-1273) x DAB1_MACFA (1-555)

QY 242 GGTGGAAGCTAAAGATGACGACGACCTGACTTCTATGCGATTCAACCTGGTGCCACT 301
 Db 57 GlyAspLysLeuCysGlnAspSerMetMetLysLeuLysGlyValAlaGlyAla 75
 QY 302 GTCCATGTCTCTCGAAAGTCTGCGCTGACCTGATCAGAACCGGACCGCTGTG 355
 Db 76 -----ArgSerLysGlyGluHisLysGlnLysLysLysLysLysLysLys 90
 QY 356 -----CACAAAGTGGCTGGCATGACGAGAGTTCGGGTGTG 391
 Db 91 PheGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 110
 QY 392 CACACTGCGCTGCAC---AGCAGCTCTCTTACAGGAGGCGCTTTTAAGATGCTCAGC 448
 Db 111 HisGluLysSerTyrIleAlaLysAspThrAspHisArgAlaPheGlyTyrAlaCys 130
 QY 449 AATAAGAGTCTCTGGATCAGATCATTTGTGGCCACCCAGCGCTCAGCAGTACCTATT 508
 Db 131 GlyLysGluGlyAsnHisArgPheValAlaLysLysLysLysLysLysLysLysLys 150
 QY 509 GCTCTTGGGGTCTCCAGCAGACGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 568
 Db 151 -----IleLeuAspLeuArgAspPheGlnLysLysLysLysLysLysLys 167
 QY 569 GATACGTTG-----GTGCTGCTCCACCCAGCGCTCTCAATGCC 607
 Db 168 GluGluLeuGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 187
 QY 608 ATTGTC-----CTGGTTCTGCATCCCGTACCA 634
 Db 188 IleLeuGluGluAspValGluAspProValTyrGlnTyrIleValPheGlu-----Ala 205
 QY 635 GSCAGTGGCCCAATG-----CCTGGAGTACTCTCTCTCCGAGGATGCCCTCCAGC 688


```
Db      394  GluAspAspGlyGluAspLeuAspThr-----ArgProSerAspLeuMet 409
QY      605  GCC-----ATTGCTGCTGCTCTCAGCTCCAGCAGGACGTGCCCAATG 649
Db      410  AlaGlnMetCysProGlnLeuValAlaThrAlaPheSerLysAspGlySerAlaThrGlu 429
QY      650  CCTGGGACTGACTCTCTCCCGAGCATGCTCCAGCTCATACGGGATATGCCAGGT 709
Db      430  AspGluAspSerAsnSerAspAspThrAspGluSerThrAspAspGluAspGluAspSer 449
QY      710  GGCTTCCTGTTTGAGGGCTCTCAGATCATATGATGAC----- 748
Db      450  AspSerGluAsnAspSerAspSerGluAspGluGluGluGluGluGluGluGlu 469
QY      748  ----- 748
Db      470  ProValLysIleAlaGlyAlaLysArgGlyArgAsnAspAspGluValSerProLeu 489
QY      749  -----TTTACCCAAACACACAGGTCCACATCTCTAGCAGTA----- 785
Db      490  AlaLysLysProArgIlePheSerProProValArg-ProArgProGlnAlaIleArgVa 509
QY      786  -----CTCCAGCTCCCGCCAGGCTCCCTGGGTACAGTGGAGCTCTGGGCCCGGCC 841
Db      509  LysLeuProSerProAlaProSerSerArgGlySerThrProValSerProValSerPr 529
QY      842  ATCACCAGAGTACGCTGGCCACGGCT-----GGCCCCGGCCAGCACT 886
Db      529  SerProLysAlaLysArgProAlaGlnAlaThrSerLeuLeuAlaSerHisProMetly 549
QY      887  CCGGAGAGCAGCTCTACACACCGACTC-----CTGGCCCGCAGGCTCATTC 934
Db      549  LysArgGluLysLeuGlnGluGluLeuArgLysAlaGlyLeuAlaProProSerAlaPr 569
QY      935  T-----CAGGACCTCACCAGTGTCTCTGTTGTCAGTGGAGCGCCCATCACCAAT 988
Db      569  ValLeuMetGlyProAspGlyValProLeuGlyThrValArgSerArgSerProSe 589
QY      989  GATCTCTTACGCCAAGCCCTACAGC-----ATGCCCTTACGGCTCTGGGCG 1036
Db      589  rValSerSerProProSerValSerValSerLeuProLeuProSerArgGlyValProSe 609
QY      1037  CCCAGCCTTCAGAGCCAGTGGAGCCCGACCTTACACAGTACGTACATGG 1088
Db      609  rGlyGlyIleLysValThrGlyAspPro-----ThrProIleValAsnArgP 624
```

Search completed: November 12, 2003 06:47:11
Job time : 41 secs

GenCore version 3.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - protein search, using frame_plus_nlp model

Run on: November 12, 2003, 06:43:14 / Search time 59.5 seconds
(without alignments)
11042.049 Million cell updates/sec

Title: us-09-735-251-3

Perfect score: 2332

Sequence: 1 GCAGAGGGCCAGAGAGAGAA.....TGCTCCATCTCTCTCTGTC 1273

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 930525 seqs, 248052604 residues

Total number of hits satisfying chosen parameters: 1463050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ nlp.model -DEV=xip
-Q=/cgn2 1/USPTO.spool/p/US9735251/runat 12112003 061635 28135/app query.fasta 1.1415
DB=SPTRMBL_23 -CPMT=fastan -SUFPIX=spc -MINMATCH=0.1 -LOGPCU=0 -LOGPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 -cd: -LIST=45
-DOCALIGN=200 -THR_SCORE=spc -THR_MAX=100 -THR_MIN=3 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=3 -MAXLEN=200000000
-USER=US9735251 -ACGN 1 1 86 -runat 12112003 061635 28135 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=C -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-GATEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELCP=6 -DELEXT=7

Database :

SPTRMBL_23:
1: sp_atchea;
2: sp_bacteria;
3: sp_fungi;
4: sp_human;
5: sp_invertebrate;
6: sp_mammal;
7: sp_mhc;
8: sp_organelle;
9: sp_plant;
10: sp_phage;
11: sp_protist;
12: sp_virus;
13: sp_vertebrate;
14: sp_unclassified;
15: sp_rviro;
16: sp_bacteriap;
17: sp_archaeap;

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	19.4	83.1	380	4	Q96103	Q96103 homo sapien

ALIGNMENTS

RESULT 1

Q96103 ID Q96103 PRELIMINARY: PRT: 380 AA.

AC Q96103; 01-DEC-2001 (TrEMBL:rel. 19, Created)
DT 01-DEC-2001 (TrEMBL:rel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBL:rel. 23, Last annotation update)
DE Similar to RIKEN CDNA 2330004C15 gene (Hypothetical protein MGCL4421).

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

RN [1] SEQUENCE FROM N.A.

RP TISSUE=uterus;

RA Strausberg R.;

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

RN [2] SEQUENCE FROM N.A.

RP TISSUE=Eye;

RA Strausberg R.;

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

RN [3] SEQUENCE FROM N.A.

RP

Q96582 homo sapien
Q91w67 mus musculus
Q9d7p5 mus musculus
Q8k141 mus musculus
Q8lsp9 oryza sativ
Q8c835 mus musculus
Q8r317 mus musculus
Q8c7t4 mus musculus
Q9nif3 dictyostella
Q9vwd9 drosophila
Q9qzm0 mus musculus
Q9sm46 arabidopsis
Q9jip9 rattus norv
Q9s119 arabidopsis
Q8nans homo sapien
Q96g51 homo sapien
Q9b6k5 homo sapien
Q8mwo7 aedes aegypt
Q29071 sus scrofa
P70433 mus musculus
Q8b117 mus musculus
Q8r411 mus musculus
Q8r6e8 homo sapien
Q8vpm8 micrococcus
Q3y076 leishmania
Q75851 homo sapien
Q9qzm1 mus musculus
Q9sm59 bos taurus
Q89037 rattus norv
Q02660 bos taurus
Q8epm4 bos taurus
Q9uhd9 homo sapien
Q9haz4 homo sapien
Q8bn13 mus musculus
Q9dh52 mycobacteri
Q9dh52 meleagrid h
Q9jig7 mus musculus
Q9qxm9 mus musculus
Q9qxm0 mus musculus
Q75046 homo sapien
Q9fwf5 oryza sativ
Q8nhd2 homo sapien
Q8bkv7 leishmania
Q99n38 rattus norv

[illegible]

Db 21 SerLeuArgLeuProGluThrGluLeuGlyGlySerLeuGlyGlyTyrSerLeu 40
 QY 164 TCATTCTGACAGCTTATGCTGGCAACTCAGGAGTCTGTTCAGACCTTGAGCTG 223
 Db 41 SerPheLeuLysGlnLeuLeuAlaGlySerGlnGluSerValProAspProGluLeu 60
 QY 224 ATTGATCTGATCTACTGTGGTGGAGCTAAAGATGACACAGACCTTGACTTGGC 283
 Db 61 IleAspLeuIleTyrCysGlyArgLysLeuLysAspGlnThrLeuAspPheTyrGly 80
 QY 284 ATTCACTGGTGCACCTGCTCAATTTCTGGGAAGTCTCTGAGCTGACCTGATCAGAA 343
 Db 81 IleGlnProGlySerThrValHisValLeuAspLysSerThrProGluProAspGlnLys 100
 QY 344 CGGAACCTGTGGCAAAAGTGGCTCCATGAGAGAGTTCGGGTGGTGCACACTGGCTG 403
 Db 101 ProGluProValAspLysValAlaAlaLeuArgGluPheArgValIleuHisThrAlaLeu 120
 QY 404 CACAGCAGCTCTCTTTACAGGAGCGGTCTTTAAGATGCTCAACATAAGAGCTCTCTG 463
 Db 121 HisSerSerSerTyrArgGluAlaValPheLysMetLeuSerAsnLysGluSerLeu 140
 QY 464 GATCAGATCATTTGTGGCCACCTCCATGAGAGTTCGGGTGGTGCACACTGGCTG 523
 Db 141 AspGlnIleIleValAlaThrProGlyLeuSerSerAspProIleAlaLeuGlyValLeu 160
 QY 524 CAGGACAGGAGCT 583
 Db 161 GlnAspLysAspLeuPheSerValPheAlaAspProAsnMetLeuAspThrLeuValPro 180
 QY 584 GGTCAACAGCCTCGTCAATGCAATTTCTGGTCTGCAATCGTACAGGAGCTGAGC 643
 Db 181 AlaHisProAlaLeuValAsnAlaIleValLeuAlaLeuHisSerValAlaGlySerAla 200
 QY 644 CCAATGCTGGAGCTGACT 703
 Db 201 ProMetProGlyThrAspSerSerSerAlaSerMetProSerSerSerTyrArgAspMet 220
 QY 704 CCAGTGGCTTCT 763
 Db 221 ProGlyGlyPheLeuPheGluGlyLeuSerAspGluAspPheHisProAsnThr 240
 QY 764 AGTCCACACCTCTAGCAGTACTTCCAGCTCTGGCCAGCTCTCTCTCTCTCTCTCTCTCT 823
 Db 241 ArgSerThrProSerSerSerThrProSerSerAlaIleValLeuSerLeuGlyTyrSerGly 260
 QY 824 GGTGCTGGGCCCCGCCATGACCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 883
 Db 261 AlaValGlyProArgProIleThrGlnSerValAlaSerAlaAlaLeuAlaSer 280
 QY 884 ACTCCGAGAGAGCTCTCACACAGCAGCTCTCTGGCAACCAAGTCTCTCTCTCTCTCTCT 943
 Db 281 ThrProGluSerSerSerHisThrProThrProThrProGlyThrGlnGlyHisSerSerGlyThr 300
 QY 944 TCACAAATGCTCTGTGGTCTCAGTCCAGGAGCGCCATCCACATGATCTCTCTCTCTCTCT 1003
 Db 301 SerProMetSerSerGlyValGlnSerGlyThrProIleThrAspLeuPheSerGln 320
 QY 1004 GGCCTACAGCATGCT 1063
 Db 321 AlaLeuGlnHisAlaLeuGlnAlaSerGlyGlnProSerLeuGlnSerGlnTyrGlnPro 340
 QY 1064 CAGTGCAGCAGCTAGTGCATGAGGCAATCCAGGAGCTGAGTGCAGCTCTCTCTCTCTCT 1122
 Db 341 GlnLeuGlnGlnLeuArgAspMetGlyIleGlnAspAspGlyLeuSerLeuArgAlaLeu 360
 QY 1123 CAGGCCACGGTGGGACATCCAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1182
 Db 361 GlnAlaThrGlyGlyAspIleGlnAlaAlaLeuGlnLeuIlePheAlaGlyValAlaPro 380

RESULT 3

Q91W67

ID Q91W67

PRELIMINARY; PRT; 380 AA.

Q91W67;
 01-DEC-2001 (TRENBLrel. 19, Created)
 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Similar to RIKEN CCNA 2300004C15 gene.
 GN 2300004C15RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 SEQUENCE FROM N.A.
 RP TISSUE=Salivary gland;
 RC Strausberg R.;
 RA Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC016456; ANH16456.1;
 DR WGD; MGJ:1916709; 2300004C15RIK.
 DR InterPro: IPR001901; SECC.
 DR InterPro: IPR000449; UBA domain.
 DR InterPro: IPR000626; Ubiquitin.
 DR Pfam: PF00627; UBA.1.
 DR Pfam: PF00240; ubiquitin.1.
 DR SMART: SM00165; UBA.1.
 DR SMART: SM00213; UBC.1.
 DR PROSITE: PS01067; SECC_SEQ61G; 1.
 DR PROSITE: PS0053; UBIQUITIN_2; 1.
 SQ SEQUENCE 380 AA; 40407 MW; 6DE0834A7CCEABF2 CRC64;
 Alignment Scores:
 Pred. No.: 5,85e-137 Length: 380
 Score: 1884.00 Matches: 371
 Percent Similarity: 98.9% Conservative: 5
 Best Local Similarity: 97.63% Mismatches: 4
 Query Match: 81.84% Indels: 1
 Gaps: 0
 DB:
 US-09-735-251-3 (1-1273) x Q91W67 (1-380)

QY 44 ATGTCTCTCTAGACTGGCAGCTGGCGGTGAAGCTGGTGCAGCCACTTACTTCCAAG 103
 Db 1 MetSerLeuSerAspTrpHisLeuAlaValLysLeuAlaAspGlnProLeuAlaProLys 20
 QY 104 TCATTCTCTGGTTCGCCAGACAGAACTGGAGAAATACTCGCTAGGGGGCTATAGTATT 163
 Db 21 SerIleLeuGlnLeuProGluThrGluLeuGlyGlyTyrSerLeuGlyGlyTyrSerIle 40
 QY 164 TCATTCTCTCAACGAGCTATTCTCTGGCAAACTCCACAGAGTCTGTTCACAGCCCTGAGCTG 223
 Db 41 SerPheLeuLysGlnLeuIleAlaGlyLysLeuGlnGluSerValProAspProGluLeu 60
 QY 224 ATTGATCTCTATCTGTGGTTCGGAAGCTTAAAGATGACACAGACTGCTTCTATGCTC 283
 Db 61 IleAspLeuIleTyrCysGlyArgLysLeuLysAspGlnThrLeuAspPheTyrGly 80
 QY 284 ATTCACTGGTGGTCCACTGCTCCATGCTTCGCAAGTCTGGCAAGTCTGGCCCTGACCTGATCAGAAA 343
 Db 81 IleGlnProGlySerThrValHisValLeuArgLysSerThrProGluProAspGlnLys 100
 QY 344 CGGAACCTGTGGCAAAAGTGGCTCCATGAGAGAGTTCGGGTGGTGCACACTGGCTG 403
 Db 101 ProGluProValAspLysValAlaAlaLeuArgGluPheArgValIleuHisThrAlaLeu 120
 QY 404 CACAGCAGCTCTCTTTACAGGAGCGGTCTTTAAGATGCTCAGCAATAGAGGCTCTCTG 463
 Db 121 HisSerSerSerTyrArgGluAlaValPheLysMetLeuSerAsnLysGluSerLeu 140
 QY 464 GATCAGATCATTTGTGGCCACCTCCATGAGAGTTCGGGTGGTGCACACTGGCTG 523
 Db 141 AspGlnIleIleValAlaThrProGlyLeuSerSerAspProIleAlaLeuGlyValLeu 160
 QY 524 CAGGACAGGAGCT 583
 Db 161 GlnAspLysAspLeuPheSerValPheAlaAspProAsnMetLeuAspThrLeuValPro 180


```

Db      221 ProGlyGlyPheLeuPheAspAlaLeuSerAspGluAspPheHisProSerThr 240
QY      764 AGTCCACACCTCTAGACAGTACTCCAGCTCCGCCAGGCTCCCTGGGTACAGTGA 823
Db      241 ArgSerThrProSerSerThrProSerArgProAlaSerLeuGlyTyrSerGly 260
QY      824 GTGTGGTGGCCCGCCCATCAGAGTGAAGTGGCAAGCTTGGCTTGGCTGGCCAG 883
Db      261 AlaAlaGlyProArgProThrGlnSerGluLeuAlaAlaLeuAlaSer 280
QY      884 ACTCCGAGCAGCAGCTCTCAGACACCGACTCTGACACCCAGGTCATCTCTAGGAGC 943
Db      281 ThrProGluSerSerSerHisHisThrProThrProGlyThrGlnGlyHisSerGlyThr 300
QY      944 TCACCAATGCTCTCTGGTCTCAGTCAGGAGAGCCCATCAGCAATGATCTCTCAGCAA 1003
Db      301 SerProMetSerSerGlyValGlnSerGlyThrProThrAspLeuPheSerGln 320
QY      1004 GCCTACAGCATGCCCTTCAGGCTCTGGGACGCCAGCTTCAGACGAGTGGCAGGCC 1063
Db      321 AlaLeuGlnHisAlaLeuGlnAlaSerGlyHisProSerLeuGlnIleGlnTrpGlnPro 340
QY      1064 CAGCTCAGCAGCTAGTACATGGGCATCCAGAGCAGTACAGCTCAGCTCGGCC-CTG 1122
Db      341 GlnLeuGlnGlnLeuArgAspMetGlyTyrGlnAspAspGlnLeuSerLeuArgAlaLeu 360
QY      1123 CAGGCCACCGGTGGGACATCCAAACAGCCCGGAGCTCATCTTTGCTGGAGGAGCCCA 1182
Db      361 GlnAlaThrGlyGlyAspIleGlnAlaAlaLeuGluLeuIlePheAlaGlyAlaPro 380

RESULT 5
ID Q8K141 PRELIMINARY; ERT; 554 AA.
AC Q8K141
DI 01-OCT-2002 (TrEMBLrel; 22, Created)
DI 01-OCT-2002 (TrEMBLrel; 22, Last sequence update)
DI 01-MAR-2003 (TrEMBLrel; 23, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/CCDS International Nucleotide
DR EMBL; BC028857; AAH28857.1;
DR InterPro; IPR000449; UBA_domain;
DR Pfam; PF00627; UBA; 1;
DR SMART; SMC0165; UBA; 1;
DR SMART; SMC0213; UBO; 1;
DR PROSITE; PS50053; UBIQUITIN_2; 1;
KW Hypothetical protein.
SQ SEQUENCE 554 AA; 58677 MW; F9255C9985A4C029 CR; 54;

Alignment Scores:
Pred. No.: 4,57e-06 Length: 554
Score: 188.50 Matches: 124
Percent Similarity: 39.74% Conservative: 75
Best Local Similarity: 22.96% Mismatches: 174
Query Match: 8.19% Indels: 99
DB: 11 Gaps: 21

US-09-735-251-3 (1-1273) x Q8K141 (1-554)
QY 86 CAGCAGCTTACTCCAAAGTCTATCTCTGCTTCCAGACAGACAGTGGGAGATCTCG 145
Db 24 GlutrolysisMetIysValThrValDysThrProLys---GluGlyGluGluPheAla 42
QY 146 CTA---GGGGGCTATAGTATTTCATTTTGAACAGCAGCTATTCTCTGGCAATCCAGGAG 202

```

```

Db      43 ValProGluAsnSerSerValGlnGlnPheLysGluGluIleSerLysArgPheLysSer 62
QY      203 TCTGTCTCCAGACCCCTGAGCTGATCTGATCTACTGTCTGCGAAGCTTAAAGATGAC 262
Db      63 HisIle-----AspGlnLeuValLeuIlePheAlaGlyGlyValLeuLysAspGln 79
QY      263 CAGACACTTGACTTCTATGGCATTCACCTGGTCCACTGCTCCATGTTCTCTCGAAG--- 319
Db      80 AspThrLeuSerGlnHisGlyIleHisaspGlyLeuThrValHisLeuValIleLysThr 99
QY      319 ----- 319
Db      100 GlnAsnArgProGlnAspAsnSerAlaGlnGlnThrAsnAlaProGlySerThrValThr 119
QY      320 TCTGTGGCTGACCTGATCAGAAACCGGAACCTGGACAAGTGGCTGCCATCAGAGAG 379
Db      120 SerSerProAlaProAspSer-----AsnProThrSerGlySerAlaAlaAsnSerSer 137
QY      380 TTCCGGGTG-----TTGCACACTGCCCCCTCCACAGCAGCTCC 415
Db      138 PheGlyValGlyLeuGlyLeuAlaGlyLeuSerSerLeuGlyLeuAsnThrThr 157
QY      416 TCTTACAGGAGCGGTCTTTAAGATG-----CTCAGCAATAAGAGTCTCTG 463
Db      158 AsnPheSerGluLeuGlnSerGlnMetGlnArgGlnLeuLeuSerAsnProGluMetMet 177
QY      464 GATCAGATCATTTGTGGCCACCCAGCCCTCAGCAGTACCCCTATTGCTCTTTGGGTTCTC 523
Db      178 ValGlnIleMet-----GluAsnProPheValGlnSerMetLeu 190
QY      524 CAGACAAGACCTCTTCTCT-----GTCTTCGCTGATCCCAATATATCTGTATAGTTG 577
Db      191 SerAspProAspLeuMetArgGlnLeuIleMetAlaAsnProGlnMet---GlnGlnLeu 209
QY      578 GTGCTCTCACCAGCCCTCTCAATGCC-----ATTGCTCTGTTCTG 622
Db      210 IleGlnArgAsnProGlnSerHisMetLeuAsnAsnProAspIleMetArgGlnThr 229
QY      623 CACTCTGTAGCAGGAGTCCCAATGCTGGAGTCTCTCTTCCCGGAGCATGCC 682
Db      230 LeuGluLeuAlaArgAsnProAlaMetMetGlnMetMetArgGlnAspArgAla 249
QY      683 TCCAGCTCATACCGGATATGCCAGTGGCTCTC-----CTGTTTGAAGGG 727
Db      250 LeuSerAsnLeuGluSerIleProGlyGlyTyrAsnAlaLeuArgArgMetTyrThrAsp 269
QY      728 CTCTCAGATGAT-----GAGGATGACTTTTACCCCAACACACAGGTCACCA 772
Db      270 IleGlnGluProMetLeuAsnAlaAlaGlnGlnPheGlyGlyAsn----- 285
QY      773 CCTCTTAGCAGTACTCCAGCTCCCGCCAGCTCTCTGGGGTACAGTGGAGCTCTCTGG 832
Db      286 ProPheAlaSerLeuValSerSerSer-----SerSerAlaGluGly 299
QY      833 CCCCAGCCCATCACCAGAGTGGCTGGCCCGCTTGGCCCTGGCCAGCTCCGAG 892
Db      300 ThrGlnPro---SerArgThrGluAsnArgAspProLeuProAsnProThrAlaProGln 318
QY      893 AGCAGCTCTCACACACCGACTCTCTGGACCCAGGCTCTCTCAGGAGCTCCACCAATG 952
Db      319 ThrSerGlnSerSerProAlaSerGlyThrThrGlySerThrThrAsnThrMetSerThr 338
QY      953 TCTCTCTGCTCCAGTCCAGGAGGCCATCACCATGATCTCTTCCAGCAAGCCCTACAG 1012
Db      339 SerGlyGlyThrAlaThrSerThrProAlaGlyGlnSerThrSerGlyProSerLeuVal 358
QY      1013 CATGCCCTTCAGGCTCT-----GGGACGCCAGCTTCAGAGCAGTGGGAG----- 1060
Db      359 ProGlyValAlaSerMetPheAsnThrProGlyMetGlnSerLeuLeuGlnIle 378
QY      1061 -----CCCAGCTGAGCAGCTACTGTCATGCGCATCCAGGAGCATGAGCTGAGC 1111

```

368	QY	GCCATGAGAGAGTTC	CGGGTG-----TTGCACACTGCCCTGCACAGCAGC	411
112	Db	AlaSerAsnSerProAlaValGlyPheGlyLeuLeuHisGlyLeuGlyGlySerGly	131	
413	QY	TCCTCTTACAGGAGCGGTCTTTAAGATGCTCAGCAATAAGGAGTCTCTGGATCAGATC	472	
132	Db	SerAlaAsnSerGlyGlyLeuGlySerPheGlySerGlyLeuProGluLeuSerGlnMet	151	
473	QY	-----ATTGTGGCACCCTCAGGAGCTCAGCAGTGCAC-----CCTATTGCT	511	
152	Db	GlnGlnGlnLeuSerGluAsnProThrLeuMetArgGluLeuMetAsnMetProLeuMet	171	
512	QY	CTTGGGTCTCCAGGACAGGAGCTCTCTCT-----GTCTTCCTGATCCCAATATG	565	
172	Db	GlnAsnIleLeuAsnSerProAspLeuIleArgAsnIleIleMetAsnAsnProGlnMet	191	
566	QY	CTTGATAGTGTGGTGGCTGTCTCACCAGACCTCGTCAATGCCATT-----GTCCTGGTT	619	
192	Db	ArgGlu-----IleValAspArgAsnProAspLeuAlaHisValLeuAsnAspProSerIle	210	
620	QY	CTGCACTCCGTAGCAGGAGTGCCTCCCAATGCCT-----GGG	655	
211	Db	LeuArgGlnThrValGluAlaAlaArgAsnProGluLeuMetArgGluMetMetArgAsn	230	
656	QY	ACTGACTCTCTCCCGAGCATGCCCTCCAGCTCATCCGGGATATCCAGTGCCTTC	715	
231	Db	ThrAspArgAlaMetSerAsnIleGluSerSer-----ProGluGlyPhe	245	
716	QY	-----CTGTTTGAAGGCTCTCAGAT-----	736	
246	Db	AsnMetLeuAspArgMetTyrGluThrValGlnGluProPheLeuAsnAlaThrThrMet	265	
737	QY	---GATGAGATGACTTTCACCCAAAC-----	760	
266	Db	AlaGlyGlyGlyAspArgSerSerAsnProPheSerAlaLeuLeuGlyAsnHisGlySer	285	
761	QY	-----ACCAGGTCCACACACCTCT	778	
286	Db	AsnGlnAlaArgAspProAlaAlaAsnSerProThrThrThrSerGluSerThrThrGly	305	
779	QY	AGCAGTACTCCAGCTCCGCCCGACCTCCCTGGGTACAGT-----	820	
306	Db	SerProAlaIleAsnThrAsnProLeuProAsnProThrSerThrAlaGlyAlaAla	325	
821	QY	---GAGCTTATCGGCGCCGCGCCATCACCCACAGTGAGCTGCCACCCCTGGCCCTG	877	
326	Db	GlnGlyAlaThrArgProSerProValThrAsnAlaArgSerAlaThrAlaGlyGlyLeu	345	
878	QY	SCCAGCATCGGAGAGACGCTTTCACACCCGACTCTCGGCCACCCAGGGTCATTCCTCA	937	
346	Db	GlyGlyLeu-----SerSerThrAspLeuGlyGlyMetLeuGlyGlyGlySer	361	
938	QY	GCGACCTCACCAATGTCCTCTGGTGTCCAGTCAGGACGCCCATCACCAATGATCTCTTC	997	
362	Db	AspThrSerPheLeuSerGlnValLeuGln-----AsnProThrMet	375	
998	QY	AGCCAGACCTTCACAGCATGCCCTT-----CAGGCCTCTGGCAG-----	1033	
376	Db	MetGlnMetMetGlnAsnIleMetSerAsnProGlnSerMetAsnGlnLeuLeuAsnIle	395	
1037	QY	---CCGACCTTCAGAGCCAGTGGCAGCCCGACCTCGACGAGTACGTACATGGGCATC	1093	
396	Db	AsnProAsnValArgAsnMetMetGluSerAsn---ThrGlnLeuArgGluMet---Phe	413	
1094	QY	CAGGACGATGATGAGCTCGCGCCCTCGACGCCACCGTGGGACATCCCAAGCAGGCC	1153	
414	Db	GlnAsnProGluPheValArgGlnLeuThrSerProGluThrLeuGlnLeuIleSer	433	
1154	QY	TGGAGCTCATCTTTGTGGAGAGAGCCCATGAACCTCCCTGCTTCGCTGAACCCCGACGA	1213	
434	Db	PheGlnGlnSerLeuMetSerGlnLeuGlyGlnGlnGlnAlaGlyProGluArgThrGln	453	

RT RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
RL EMBL; AK049298; PAC33665.1;
SC SEQUENCE 555 AA; 59497 MW; 7275E3735A03937E CRC64;

EMBL; AK049298; EAC33666.1; S.
SEQUENCE 555 AA: 59497 MW: 7275E3735A03937E CRC64:


```

Dh 77 ProPro-----ProProProValGluGlnGlnValProThrProSerAsnThrGln 93
Qy 362 -----G-TG 364
Db 94 ProGlnGlyLeuProGlyValProGlnAsnLeuAsnAspMetMetAsnProMetLeu 113
Qy 365 GGTGGCATGAGAGAGTTCGGGTGTGACAATCCCTCTCAGACAGAGTTCCTTACAG 424
Db 114 GlnGluMetPheAsnSerArgMetAspSerLeuLeuAspAsnProAspPheArg 133
Qy 425 GAGCGGTCTT-----AGATGCTCAGCAAAAGAGAGTCTCTGAT 466
Db 134 AspMetMetGlyAsnProGluMetArgGluValLeuAsnAsnProGluValAla 153
Qy 467 CAGATCATGTGGCCACCCAGCGCTCAGCATGACCT-----ATTGCTCT 514
Db 154 GlnMetLeu-----SerAspProArgGlnLeuArgGlnSerLeu 166
Qy 515 GGGTTCCTCAGACAGAGACCTCTCTCTTTC-----GCTATCCCATATG 565
Db 167 GluMetMetArgAsnProGluLeuMetArgGluMetMetArgAsnAlaAspArgAlaMet 186
Qy 566 CTGTGATACGTGGTGGCTGCTGCTGCA-----592
Db 197 IleAsn-----IleGAsnSerProGluGlyPheAsnLeuLeuArgMetTyrThr 204
Qy 593 -----GCCCTGCTGATGGATG-----GTCTGCTCTGCTGACTCGGTACCA 634
Db 205 AspLeuGlnGluProLeuMetAsnAlaAlaAsnGlnAlaAlaSerGlnAsnGlnThr 224
Qy 635 GGCAGTGGCCCAATG--CTGGGACTCAATTTCTCCGAGAGTGCCTCCAGCTCA 691
Db 225 AsnSerAsnProIleGlnThrAsnThrAspAlaAsnAspAsnSerGlnPro-----241
Qy 692 TACCGGATATGCCAGTGGCTCTCTCTCTGAGGCTCTGATGATGATGATGATGAT 751
Db 242 --LeuProAsnProThrSerThrAsnSerSerSer-----252
Qy 752 CACCCAAACACACAGGTGCACACCTCTAGCACTATCCCGAGCTCCGGCCA-----GCC 805
Db 253 -----ThrSerSerAsnProThrSerSerSerProSerSerArgProThrThrGly 269
Qy 806 TCGTGGGTACATGAGTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 865
Db 270 SerSerThrAsnThrArgAlaAsnAsnAlaAlaAsnThrPheSerThrGlyGly 289
Qy 866 GTTGGCGCTGCCAGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 925
Db 290 GlyMetGlyGlyGlyThrAsnAsnThrGlyThrAsnAsnThrGlySerThrAsnAsnThr 309
Qy 926 GGTCACTCTCAGGAGCTCAGCAATTCCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 979
Db 310 GlyAlaSerAsnProThrAlaSerMetPheGlyGlyGlyGlyGlyGlyGlyGlyMet 329
Qy 980 -----ATCACCAT 988
Db 330 GlyGlyMetGlyGlyMetLeuGlyMetAspProGluValGlnGlnLeuLeuAsnAsn 349
Qy 989 GATCTCTCAGGACAGCTTACAGATGAGTCTGAGAGCTTGGCAGCAGCCAGCTTAG 1048
Db 350 ProValAlaGlnGlnMetMetGlnArgGluMet-----SerAspProAlaMetMet 366
Qy 1049 AGGCAG-----TGGCAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1099
Db 367 GlnGlnMetLeuThrMetAsnProGlnLeuValGlnMetMetValAspSerAsnProGlnLeu 386
Qy 1100 GATGAGCTGAGCTGGCGCT-----1120
Db 387 ArgGluAlaMetAsnAsnPro 393

```

RESULT 11

Q9VWD9

10 Q9VWD9

PRELIMINARY;

PRT: 547 AA.

```

AC Q9VWD9;
DI 01-MAY-2000 (TrEMBLrel. 13, Created)
DI 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DI 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG14224 protein [L38919p].
GN CG14224.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=2019606; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov V.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evansgelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C.A., Gabrielian A.E., Garg N.S., Geibart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Calalai M., Klaus P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matte B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinet K.C., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
RA Sheth B.C., Siden-Kiamos J., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton X., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungai C.J., Nunoo J., Pacleb J., Paragass V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003522; AA049003.1; -
DR EMBL; AY128447; AA075040.1; -
DR HSSP; P02248; 17H.
DR FlyBase; FB0001057; CG14224.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000426; Ubiquitin.
DR Pfam; PF00627; UBA, 1.
DR Pfam; PF00240; Ubiquitin; 1.
DR PROSITE; PS00553; UBIQUITIN 2; 1.
DR SEQUENCE 547 AA, 58836 MW, 8496DC6EC165C4D7 CRC64;

```

Db		301	LeuProPrcAsnAsnValLeuAsnThrProAlaMetArgSerLeuLeuGlnMetAla	320
Qy		778		778
Db		321	AspAsnFroAlaMetMetGlnAsnLeuLeuAsnAlaProTyrThrArgSerMetMetGlu	340
Qy		779		779
Db		341	SerMetSerGlnAspProAspMetAlaAlaArgLeuSerSerProLeuMetSer	360
Qy		794	TCCCGGCCACCTCCCTGGGGTACAGTGCGAGCTGTGGCGCCCGCCATCATCCCAGAGT	853
Db		361	AsnAsnProAlaLeuGlnGluInValArgGlnMetMetProGlnPheMetAlaGlnMet	380
Qy		854	GAGCTGGCCACCCTTTGGCCCTGGCCAGCATTCCGAGAGCAGCTCT	901
Db		381	GlnAsnProGluValMetAsnMetLeuThrAsnProAspAlaMetAsnAlaIleLeuGln	400
Qy		902		919
Db		401	IleGlnGlyMetGluGlnLeuArgSerAlaAlaProGlyLeuValGlyThrLeuGly	420
Qy		920		952
Db		421	IleProProProProGlyAlaGlyThrGlyThrAsnProAlaSerGlyAspGlySer	440
Qy		953		973
Db		441	GlyGlyAsnSerGlyAlaSerThrAsnAsnValSerProSerSerGlyLeuAsnAlaGly	460
Qy		974	ACGCCATCACCAAT	1006
Db		461	ThrGlyThrProAsnLeuAlaProGlyG.Y.GlyProAsnAlaGlnLeuPheAsnAspPhe	480
Qy		1007	CTACASCATGCCCTTCAGGCC	1048
Db		481	MetMetArgMetLeuAsnGlyMetSerAsnAsnAlaAspAsnThrGlnProPro	499
Qy		1049	AGCCATGGGCAGCCCGCCATGGCAGCAGCTACGTGACATGGGCATCCAGGACGATGAGCTG	1108
Db		500	ValArgTyGlnSerGlnLeuGluGlnLeuAsnAlaMetGlyPheValAsnArgAspAla	519
Qy		1109	AGCCTCG-CCCTCGAGGCCACCGTGGGGACATCAACGAGCCCTGGAGCTCATCTTT	1167
Db		520	AsnLeuGlnAlaLeuLeAlaThrPeg.VASPLeAsnAlalaValGluArgLeuLeu	539
Qy		1168	GCT	1177
Db		540	Ser	540
<hr/>				
RESULT 12				
Q9QZM0 PRELIMINARY; PRT; 638 AA.				
ID	Q9QZM0			
AC	Q9QZM0			
DT	01-MAY-2000 (T-EMBLrel_13, Created)			
DT	01-MAY-2000 (T-EMBLrel_13, Last sequence update)			
DT	01-MAR-2003 (T-EMBLrel_23, Last annotation update)			
DE	PLIC-2.			
GN	UBQLN2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa;			
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	NCBI_TaxID=10090;			
RC	SEQUENCE FROM N.A.			
RA	STRAIN=Swiss Webster/NIH;			
RT	Wu A.-L., Wang J., Zheleznyak A., Brown E.J.;			
RL	"Ubiquitin-related proteins regulate interaction of vimentin			
DR	intermediate filaments with the plasma membrane.";			
DR	Mol. Cell 0:0-0(1999).			
DR	EMBL; AF177346; AAF01366.1; ..			
DR	HSSP; PC2248; ITBE.			
DR	MGD; MGI:1860283; Ubqln2.			

QY	716	-----CGTTTGAAGGGCTCTCAGATGAT-----	-----GAGGATGACCTTT	751
Db	279	LeuARGAIGMetTYrThrASpIleGIingLIPrometLeuAsnAlaAglagInGluInPhe		298
QY	752	CACCCAAACACCGGTCCACACCCCTCTAGCAGTACTCCGAGCTCCGCCGAGCCTCCCTG		811
Db	299	GlyGlyAsnPropheAaThrValGlySerSerThrSergly-----		313
QY	812	GGGTACAGTGAAGCTGTGGCGCCCGCCCATCACCCAGAGTGAAGCTGGCCACCGCTTG		871

Alignment Scores:			
Seq. ID	Score	Length	Align
80	166.50	103	103
24	161.44	103	103
131	161.44	103	103
43	161.44	103	103
188	161.44	103	103
63	161.44	103	103

Db 428 proLeuPheThrSerAsnProGlnLeuGlnMetArgProGlnLeuProAsnPhe 447
QY 1169 CTGGAGAG-----CCCCATGAATCCCTCCCTGCTCCCTGAACCCCGACGAGTTGC 121
Db 448 LeuGlnGlnMetGlnAsnProGluThrIleAlaAlaMetSerAsnProArgAla----- 465
QY 1220 AGAGGCTACTGCCCTTGGGAGGCACATCATGAGGTCCTCCATCTCTC 1267
Db 466 -----MetGlnAlaLeuMetGlnIleGlnGlnClyLeu 476

RESULT 13
Q9SN46
ID Q9SN46 PRELIMINARY; PRT; 839 AA.
AC Q9SN46
DT 01-MAY-2000 (EMBLrel. 13, Created)
DT 01-MAY-2000 (EMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (EMBLrel. 23, Last annotation update)
DE Extensin-like protein.
GN F28A21.80 OR At4G18670.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Mueller M.W., Muendlein A., Feiber R., Bancroft I.,
RA Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.,
PL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mueller M.W., Muendlein A., Feiber R., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]

QY 248 AAGCTAAAGATGACACACATCTGATCTATGACATCAACTGCTGCTCCTCAT 307
Db 80 IleLeuLysAspGlnAspThrLeuMetGlnHisGlyIleHisAspGlyLeuThrValHis 99
QY 338 GTTCTGCCAAGATCTCTGG-----CCTGAATGTATAGAAACCGGACCTGTGTGACAAA 361
Db 190 LeuValIleLysSerGlnAsnAspProGlnGlyGlnAlaThrThrGlnProSerThrThr 119
QY 362 GTGCTGCCATGAGAGCACTTGGGAGGTTGTACACTGCG----- 400
Db 120 AlaGlyThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 139
QY 431 ----- 450 ATGATCTCTTCTTAAAGGAG 427
Db 140 SerSerAlaProArgSerSerSerSerSerSerSerSerSerSerSerPheClyLeu 159
QY 428 GCGGTCTTATAGATCTCAGCAAT----- 454
Db 160 GlySerLeuSerSerLeuSerAsnLeuGlyLeuAsnSerPheAsnPheThrGluLeuArg 179
QY 455 GAGTCTCTGGATCAGATCATGTGCGCCACCTCA-----GGCTCAGCAGTGAC 502
Db 180 AsnGlnMetGlnGlnGlnLeuLeuAlaSerProGlnMetMetIleGlnIleMetGluAsn 199
QY 503 CCTATTGCTCTGGGGTCTCTCAGACAGAGACCTCTCTCT-----GTCTTCTCTGAT 556
Db 200 ProPheValGlnSerMetLeuSerAsnProAspLeuMetArgGlnLeuMetAlaAsn 219
QY 557 CCCAATATGCTTGATACGTGTGTGCTCTCACCAGACCTCTCGTCAATGCC----- 607
Db 220 ProGlnMet---GlnGlnIleIleGlnArgAsnProGluIleSerHisLeuLeuAsnAsn 238
QY 608 -----ATTGCTCTGGTCTTGCACTCCGTAGCAGGACGTGCCCAATGCTCGGACTGAC 661
Db 239 ProAspIleMetArgGlnThrLeuGluIleAlaArgAsnProAlaValMetMetGlnGluMet 258
QY 662 TCTCTCTCCGAGGATCCCTCTCAGCTCATACCGGGATATGCCAGGTGGCTTC----- 715
Db 259 MetArgAsnGlnAspLeuAlaLeuSerAsnLeuGluSerIleProGlyGlyTyrAsnAla 278

RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

CR EMBL; AL035526; CAB37452.1; ...
 DR EMBL; AL161549; CAB78869.1; ...
 DR InterPro; IPR002951; Atrophin.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007290; LRR_P1ant
 DR InterPro; IPR002965; P-rich_exten.
 CR Pfam; PF00560; LRR; 4.
 DR PRINTS; PR01222; ATROPHIN.
 DR PRINTS; PR01217; PRICHEXTENK.
 DR PROSITE; PS50502; LRR_PS; 3.
 SQ SEQUENCE 839 AA: 90243 MW: 35046 F422429759 CRC64;

Alignment Scores:
 Pred. No.: 0.00306 Length: 839
 Score: 165.50 Matches: 97
 Percent Similarity: 34.86% Conservative: 25
 Best Local Similarity: 27.71% Mismatches: 121
 Query Match: 7.13% Indels: 107
 DB: 10 Gaps: 14

US-09-735-251-3 (1-1273) x Q9SN46 (1-839)

QY 312 TCGAAAGTCTGGCCCTG...AACCTGATCAGAAACCGGAACCTGTGCAAAAGTGGCTG 368
 DB 337 CysGlnLeuProArgLeuGluAsnPheThrPheSerTyraSerPheThrGly----- 354
 QY 369 CCAATAGAGATTTCGGGTGTGGACATGGCT-----TUCACA 407
 DB 355 -----GluProValCysLeuGlyLeuProGlyPheAspArgATGASnCysLeu 372
 QY 458 CGAGCTCTCTTACAGGAGGAGGCTTTAAGATGCTCAGCAATAAGAGTCTGTGATC 467
 DB 373 ProAaArgProAlaGlnArgSerProGlyCysAlaAlaPheSerSerLeuProPr 392
 QY 468 AGATCATTTGGCCAC-----GCCAG 488
 DB 392 eValAspCysGlySerPheGlyCysGlyArgSerThiArgProValValProSe 412
 QY 489 GCCTCAGAGTACGCTTATCTCTGGGTCTCCAGGACAGGACCTCTTCTGTCT 548
 DB 412 rProProThrPro-----SerProGlyGlySerPro----- 423
 QY 549 TCGCTATCCCAATATGCTTATATAGTTTGGTACAGGCTTGGTAAATGCCA 609
 DB 424 -----SerProSerLeuSerPro----- 430
 QY 609 TTGCTCTGGTCTGCACTCCCTAGAGAGTACAGAGTACAGAGTACAGTCTCTT 658
 DB 430 oSerPro-Proile-----ThiValProSerProProThr-ThrProSer 444
 QY 669 CCGGAGCATGCCCTCCAGCTCAGAGGAGATATGCGGAGTGTCTCTTTGAGGGG 728
 DB 445 ProGlyGlySerProProSerProSerSerLeuVal----- 455
 QY 729 TCTCAGATGATGAGATGACTTTACCAACACACAGGTCCACACCTCAGCAGTCTC 788
 DB 456 -----ProSerProProSerThrThr 462
 QY 789 CCAGCTCCGCGCCACCTCCCTGGGTGACAGTACAGTCTGGTGGCGGCGCCATCACCC 848
 DB 463 ProSerProGlySerProProProThrThrProThr-----ProGlyGlySerPro 481
 QY 849 AGAGTGAAGTGGCCACCGCTGGCCCTGGCCAGACCTCCGAGAGAGTCTCTCACAC 908
 DB 482 ProSerSerProThrThrProThrPro-GlyGlySerProProSerSerProThrThrP 501
 QY 909 CGATCTCTGGACCCAGGCTATTCTCTCAGGAGTCTACCAATGCTCTGTGCTCCAGT 969
 DB 501 roThrProGlyGlySerProProSerSerProThrThrProSerProGlyGly-SerPro 520

QY 969 CAGGAGCGCCCATCAATGATCTCTTACGCAAGCCCTACAGCATGCCCTTCAAGCCT 1028
 DB 521 -----ProSerProSerLeuSerProSerProProThrValPro-----SerPro 536
 QY 1029 CTGGGACGCGCCAGCTTTCAGAGCAGTGGCAGCCCGAGCTGCGAGCATGTCATG 1088
 DB 537 ProSerThrProThr-----SerProGlySerProProSerProSerProThrPro 554
 QY 1089 GCATCCAGGACGATGAGCTGAGCTGCGGCGCTGCGAGCCCGGCGGACATCCAGC 1148
 DB 555 SerSerProLeProSerProProThrProSerThrProProThrProThrProThr 574
 QY 1149 AGCCCTGAGCTCATCTTTGCTGCGAGGAGGCC-----CATGAACCTCCCTGCTT 1196
 DB 574 LYGnAsnSerProProleleProSerProProPheThrGlyProSerProProSerS 594
 QY 1197 CCCTGAGACCCCGAGGAGTGGAGAGGCTACTGCCCTTGGGAGGACCTCATGAGGTGC 1256
 DB 594 erProSerProPro-----Leup 600
 QY 1257 CTCGATCTCTCCCT 1270
 DB 600 roProVallePro 604

RESULT 14

Q9JJP9 PRELIMINARY; PRT; 582 AA.
 AC Q9JJP9;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE DA41.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Ozaki T., Hishiki T., Nakagawa A., Sakiyama S.;
 RT "Identification of a new protein that can associate with DAN.";
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D87950; BAA92267.1; ...
 DR HSSP; P02248; ITBE.
 DR InterPro; IPR006636; STIL.
 DR InterPro; IPR00449; UBA_domain.
 DR InterPro; IPR006626; Ubiquitin.
 DR Pfam; PF00627; UBA; 1.
 DR Pfam; PF00240; ubiquitin; 1.
 DR SMART; SM00727; STIL; 4.
 DR SMART; SM00165; UBA; 1.
 DR SMART; SM00213; UBA; 1.
 DR PROSITE; PS50053; UBIQUITIN 2; 1.
 SQ SEQUENCE 582 AA: 62071 MW: 665159509071E5B CRC64;

Alignment Scores:
 Pred. No.: 0.000301 Length: 582
 Score: 165.00 Matches: 117
 Percent Similarity: 32.70% Conservative: 73
 Best Local Similarity: 20.14% Mismatches: 153
 Query Match: 7.17% Indels: 239
 DB: 11 Gaps: 24

US-09-735-251-3 (1-1273) x Q9JJP9 (1-582)

QY 86 CAGCCACTTACTCCAAAGTCTTCTTGGTTCGACAGACAGAACTGGGAGATCTCG 145
 DB 24 GluProLysLeuMetLeuValThrValLysThrProLys---GluLysGluGluPheAla 42
 QY 146 CTA---GGGGCTATAGTATTTCCTTTCAGAGCAGCTTATCTGCGCAAACTCCAGGAG 202
 DB 43 ValProGluAsnSerSerValGlnGlnPheLysGluLysLeuSerLysArgPheLysSer 62

Search completed: November 12, 2003, 06:49:21
Job time : 75.5 secs

```

Alignment Scores:
Pred. No.: 0.000322 Length: 538
Score: 164.50 Matches: 99
Percent Similarity: 35.88% Conservative: 40
Best Local Similarity: 23.94% Mismatch: 140
Query Match: 7.15% Trunc: 1
DB: 12 Gap: 18

US-09-735-251-3 (1-1273) x Q9S117 (1-534)
QY 158 AGATTTCATTTCTGACGACGCTTATCTGTTGTAAGATTCAGTAGCTCTTTCGAGACCT 217
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39 ThrValGluSerPheLysGLeuLeaL...GlnAsnSerAspValProAla 55
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 218 GAGCTGATTGATCTCATCTAGCTGGTCAAGGTAAGAAGATACACAGACACTTGACT 277
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 56 AsnGlnGlnArgLeuIleTyrLysGlyArgIleLeuLysAspAspGlnThrLeuLeuSer 75
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 278 TATGGCATTCAACCTGGGTCACCTGCTCATCTTTTGGGAAGGCTCGGCTGAACTGAT 337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76 TyrGlyLeuGlnAAspHisThrValHisMetValArgGlyPheValProSerSerPro 95
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 338 CAGAAACCG... 346
      |||||
Db 96 SerAlaProAlaAlaAsnAlaGlyAsnGlnThrThrAlaProGlnAlaValLysSerAs 115
      |||||
QY 347 .....GAACCTGTG 355
      |||||
Db 116 AspSerSerAsnLeuGlyGlyGlyGluSerLeuIleProGlyLeuGlyPheAsnProLeu 135
      |||||
QY 156 GACAAAGTGGCTGCCATGAGAGAGTTC... 388
      |||||

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_rnp model

Run on: November 12, 2003, 06:19:14 ; Search time 46 Seconds
Without alignment:
8785.171 Million cell updates/sec

Title: US-09-735-251-3

Perfect score: 2302

Sequence: 1 GCGAGAGCGGAGAGAGAGAA... TGGCTTCATCTCCCTGTC 127;

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delcp 6.0 , Delcext 7.0

Searched: 1107863 seqs, 158726577 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_rnp_model -DEV=xlp
-C/cgn2.1/USPTO_speci_p/US09735251/runat_12112003_061634_28110/app_query.fasta_1.1415
-DB=A Geneseq_19Jun01 -QWTF=fastan -SUFFIX=rag -MINMATCH=0.1 -LCOPE=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -VATRLX=blosum62 -TRANS=human45odi
-S=45 -DOALIGN=200 -THR SCORE=100 -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFN=pco -NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09735251 -CGEN 1 1.81 -grunat_12112003_061634_28110 -NCPU=6 -CPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPRACK=100 -LCLNG=0
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELCP=6 -DELCEXT=7

Database : A Geneseq_19Jun01

1: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1981.DAT
2: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1982.DAT
3: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1983.DAT
4: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1984.DAT
5: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1985.DAT
6: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1986.DAT
7: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1987.DAT
8: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1988.DAT
9: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1989.DAT
10: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1990.DAT
11: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1991.DAT
12: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1992.DAT
13: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1993.DAT
14: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1994.DAT
15: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1995.DAT
16: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1996.DAT
17: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1997.DAT
18: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1998.DAT
19: /SIDSI/cgdata/geneseq/geneseq_emb1/AA1999.DAT
20: /SIDSI/cgdata/geneseq/geneseq_emb1/AA2000.DAT
21: /SIDSI/cgdata/geneseq/geneseq_emb1/AA2001.DAT
22: /SIDSI/cgdata/geneseq/geneseq_emb1/AA2002.DAT
23: /SIDSI/cgdata/geneseq/geneseq_emb1/AA2003.DAT
24: /SIDSI/cgdata/geneseq/geneseq_emb1/AA2004.DAT

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2133	92.7	410	18	AAW15414
2	2129	92.5	426	18	AAW15413
3	1959	85.1	420	20	AAV42382
4	1959	85.1	420	22	AAJ39074
5	1959	85.1	420	23	ABW56783
6	1908	82.9	380	22	ABW56783
7	1686	73.2	341	23	ABW56783
8	94.1	40.9	191	22	AAW63948
9	757	32.9	164	22	AAW63948
10	620	26.9	120	22	AAW63950
11	610	26.5	128	22	AAO13489
12	483	20.5	110	22	AAW63951
13	444	19.3	114	21	AAW58780
14	389	16.9	76	21	AAW58780
15	172.5	7.5	547	22	ABW69125
16	170.5	7.4	329	23	ABJ10942
17	164.5	7.1	536	21	AAW30040
18	164.5	7.1	577	21	AAW30039
19	162	7.0	693	23	ABW69529
20	161.5	7.0	19938	24	ABW76678
21	159	6.9	783	19	AAW37151
22	159	6.9	787	19	AAW37152
23	159	6.9	802	19	AAW37153
24	159	6.9	802	22	AAU09139
25	158.5	6.9	19938	24	ABW76679
26	157.5	6.8	245	22	ABG19161
27	156.5	6.8	617	23	AAE01355
28	156.5	6.8	617	23	ABG79641
29	156.5	6.8	617	23	ABW81074
30	156.5	6.8	617	23	ABW81074
31	156	6.8	598	22	ABG14000
32	152.5	6.6	441	22	ABW11413
33	152.5	6.6	572	18	AAW31855
34	152.5	6.6	624	22	AAW47122
35	152.5	6.6	624	23	ABW97373
36	152.5	6.6	763	18	AAW31852
37	152	6.4	19938	24	ABW76679
38	152	6.4	1040	22	ABG14734
39	150.5	6.5	819	22	ABG03569
40	149	6.5	371	22	ABG03569
41	149	6.5	415	22	ABW59727
42	149	6.5	616	23	ABW81078
43	148	6.4	1023	23	AAU82954
44	147.5	6.4	489	23	ABG4061
45	147.5	6.4	505	22	AAE01380

ALIGNMENTS

RESULT 1
AAW15414
ID AAW5414 standard; Protein; 410 AA.
XX
AC AAW15414;
XX
DT 06-JUL-1997 (first entry)
XX
DE Activated platelet protein-2 APP-2 alternatively spliced variant.
XX
KW APP-2; activated platelet protein-2; thrombus; monoclonal antibody.
XX
OS Homo sapiens.
XX
PN W09712898-A1.
XX
PD 10-APR-1997.
XX

PF 04-OCT-1996; 96WO-US15922.
 XX
 PR 06-OCT-1995; 95US-0005074.
 XX
 PA (HARD) HARVARD COLLEGE.
 XX
 PI Clement CY, Reed G;
 XX
 DR WPI; 1997-226153/20.
 XX
 DR N-FSDB; AAT64550.
 XX
 PT Naturally-occurring platelet activation polypeptide, APP-2 - used
 PT for locating a platelet thrombus in an animal, or activated platelet
 PT complex in a biological sample
 XX
 PS Claim 2; Fig 6; 50pp; English.
 XX
 CC A polypeptide (AAW15414) comprises an alternatively spliced variant
 CC of novel human activated platelet protein-2 (APP-2) (see also
 CC AAW15413), a protein that is preferentially expressed on activated
 CC human platelets but not resting platelets. Its amino acid sequence
 CC was deduced from an isolated cDNA clone (AAT64550). APP-2 is
 CC characterised as contg. an epitope that binds to monoclonal antibody
 CC 3B2 (ATCC CRU 1986). APP-2 can be produced in transformed host
 CC cells. Anti-APP-2 antibodies are useful in methods for detecting
 CC activated platelets in biological samples, for localising a platelet
 CC thrombus in an animal, and for detecting an activated platelet
 CC protein complex in an animal.
 XX
 SQ Sequence 410 AA;

Alignment Scores:
 Pred. No.: 6, 16e-177 Length: 410
 Score: 2133.00 Matches: 410
 Percent Similarity: 100.00% Conservative: 3
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 92.66% Indels: 0
 DB: 18 Gaps: 0

US-09-735-251-3 (1-1273) x AAW15414 (1-410)

QY 44 ATGCTCTCTCAGAGTGGGACCTTCGCTGAAAGTGGTGGAGCAGGACCTTACTCCAAAG 103
 Db 1 MetSerLeuSerAspThrPheLeuAlaValGlyLeuAlaAspGluProLeuThrProLys 20
 QY 104 TCTATGCTCGGTGGGAGAAAGAAATGTAAGAAATGTAAGAAATGTAAGAAATGTAAG 163
 Db 21 SerLeuArgLeuProGluThrGluLeuGlyLeuGlyLeuGlyLeuGlyLeuGlyLeu 40
 QY 164 TCATTTCTGAAGCAGCTTATGCTGCTGAAAGTGGTGGAGTGGTGGAGTGGTGGAGTGG 223
 Db 41 SerPheLeuLysGlnLeuLeuAlaGlyLeuLysLeuLysLeuLysLeuLysLeuLys 60
 QY 224 ATTGATCTGATCTACTGCTGGGAGAGCTAAAGATGACAGACACTTACTTCTATGGC 283
 Db 61 IleAspLeuIleThrCysGlyArgGlyLeuLysAspGlnThrLeuAspPheThrGly 80
 QY 284 ATTCAGCTGGTCCACTGCTCAATGCTCTGCAAGTGGTGGAGTGGTGGAGTGGTGGAG 343
 Db 81 IleGlnProGlySerThrValAlaMetArgGluPheArgValLeuHisThrAlaLeu 100
 QY 344 CCGAAGCTGTGGACAAAGTGGTGGAGTGGTGGAGTGGTGGAGTGGTGGAGTGGTGGAG 403
 Db 101 ProGluProValAspLysValAlaAlaMetArgGluPheArgValLeuHisThrAlaLeu 120
 QY 404 CACAGAGCTCTCTTACAGAGAGGCGGTCTTAAAGATGCTCAGCAATAGGAGTCTCTG 463
 Db 121 HisSerSerSerSerThrValAlaAlaMetArgGluPheArgValLeuHisThrAlaLeu 140
 QY 464 GATCAGATCTATGTCGACCCAGGCTCAGAGTGGTGGAGTGGTGGAGTGGTGGAGTGG 523
 Db 141 AspGlnIleValAlaThrProGlyLeuSerSerAspProIleAlaLeuGlyValLeu 160

QY 524 CAGCACAAGGACCTCTCTCTGCTCTGCTGATCCCAATATGCTGATGATGCTGCTGCT 583
 Db 161 GlnAspLysAspLeuPheSerValPheAlaAspProAsnMetLeuAspThrLeuValPro 180
 QY 584 GCTCACCAGCCCTCGTCAATGCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 643
 Db 181 AlaHisProAlaLeuValAsnAlaIleValLeuValLeuHisSerValAlaGlySerAla 200
 QY 644 CCAATGCTGGGACTGACTCTCTCTCCGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTG 703
 Db 201 ProMetProGlyThrAspSerSerArgSerMetProSerSerSerSerSerSerSer 220
 QY 704 CCAGTGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 763
 Db 221 ProGlyGlyPheLeuPheGluGlyLeuSerAspGluAspPheHisProAsnThr 240
 QY 764 AGTCTCACACCTCTAGCAGTACTCTCCAGTCTCCGCTGCTGCTGCTGCTGCTGCTGCT 823
 Db 241 ArgSerThrProSerSerSerThrProSerSerArgProAlaSerLeuGlyThrSerGly 260
 QY 824 GCTGCTGGGCGCCGCTCACCACAGTGGTGGCAGTGGTGGCAGTGGTGGCAGTGGTGG 883
 Db 261 AlaAlaGlyProArgProIleThrGlnSerGluLeuAlaThrAlaLeuAlaLeuAlaSer 280
 QY 884 ACTCCGAGAGCAGCTCTCACACACCCAGTGGTGGCAGTGGTGGCAGTGGTGGCAGTGG 943
 Db 281 ThrProGlySerSerSerSerHisThrProThrProGlyThrGlnGlyHisSerSerGlyThr 300
 QY 944 TCACCAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1003
 Db 301 SerProMetSerSerGlyValGlnSerGlyThrProIleThrAsnAspLeuPheSerGln 320
 QY 1004 GCCCTACAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1063
 Db 321 AlaLeuGlnHisAlaLeuGlnAlaSerGlyGlnProSerLeuGlnSerGlnThrProGln 340
 QY 1064 CAGTGGCAGCAGCTACGTGATGCGGATCCAGGATCCAGGATCCAGGATCCAGGATCCAG 1123
 Db 341 GlnLeuGlnGlnLeuArgAspMetGlyIleGlnAspAspGluLeuSerLeuArgProCys 360
 QY 1124 AGGCAACGCTGGGGAGATCCAGCAGCCTGAGCAGCCTGAGCAGCCTGAGCAGCCTGAG 1183
 Db 361 ArgProProValGlyThrSerLysGlnProThrProSerSerSerSerLeuLeuGluProHis 380
 QY 1184 GAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1243
 Db 381 GlnLeuProAlaSerProGluProProAlaSerCysArgGlyThrCysProThrProGluAla 400
 QY 1244 CTCATGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1273
 Db 401 LeuMetLysValProProSerLeuProVal 410

RESULT 2
 AAW15413
 ID AAW15413 standard; Protein; 426 AA.
 XX
 AC AAW15413;
 XX
 DT 06-JUL-1997 (first entry)
 XX
 DE Human activated platelet protein-2 APP-2.
 XX
 KW APP-2; activated platelet protein-2; thrombus; monoclonal antibody.
 XX
 OS Homo sapiens.
 XX
 PN WO9712898-A1.
 XX
 PD 10-APR-1997.
 XX
 PF 04-OCT-1996; 96WO-US15922.
 XX
 PR 06-OCT-1995; 95US-0005074.

XX (HARD) HARVARD COLLEGE.

XX Clement CV, Read G;

XX WPI: 1997-226153/20.

XX N-PSDB: AAT64549.

XX Naturally-occurring platelet activation polypeptide, APP-2 - used
 for locating a platelet thrombus in an animal, or activated platelet
 complex in a biological sample

XX Claim 1; Fig 5; 50pp; English.

XX A novel 25 kDa polypeptide (AAW15413) designated activated platelet
 protein-2 (APP-2) is preferentially expressed on activated human
 platelets but not resting platelets. APP-2 is characterized as
 contg. an epitope that binds to monoclonal antibody 3B2 (ATCC CRL
 11986). Its amino acid sequence was deduced from a cDNA clone
 (AA64549) obtd. from a human bone marrow cDNA library. An
 alternatively spliced variant (see also AAW15414) was also isolated.
 CC APP-2 can be produced in transformed host cells. Anti-APP-2
 CC antibodies are useful in methods for detecting activated platelets
 CC in biological samples, for localizing a platelet thrombus in an
 CC animal, and for detecting an activated platelet protein complex in
 CC an animal.

XX SQ Sequence 426 AA;

Alignment Scores:
 Pred. No.: 1,39e-176 Length: 426
 Score: 2129.00 Matches: 494
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 92.48% Indels: 0
 DB: 18 Gaps: 0

US-09-735-251-3 (1-1273) x AAW15413 (1-426)

QY 44 ATGTCTCTCAGACTGGCAGCTGGCGTGAAGTGGCTGACAGCCACTTACTCCAAAG 103
 DB 1 MetSerLeuSerAspTTPHisLeuAlaValLysLeuAlaAspGlnProLeuThrProLys 20
 QY 104 TCTATTCTCGTGGCAGACACAAATGGGAATACTAGCTAGGGGCTATAGTATT 163
 DB 21 SerileLeuArgLeuProGlnThrHisLeuGlyTyrSerLeuGlyTyrSerile 40
 QY 164 TCATTCTGAGCAACTATTCTGGAGAAATGAGAGTCTTCCAGACCTGAGCTG 223
 DB 41 SerPheLeuLysGlnLeulleAlaGlyLysLeuThrHisLeuSerAlaProAspProGluLeu 60
 QY 224 ATTGATCTGATCTACTGTGGGAAGCTTAAAGATGACACAGACACTTACTTATGCC 283
 DB 61 IleAspLeulleTyrCysGlyArgLysLeuLysAspGlnThrLeuAspPheTyrGly 90
 QY 284 ATTCAAGCTGGTCCACTGCTCATGTTCTGCGAAGTCTGGGCTGAACCTGATCAGAAA 343
 DB 81 IleGlnProGlySerThrValHisValLeuArgLysSerTTPProGluProAspGlnLys 100
 QY 344 CCGAAGCTGTGGCAAGTGGCTGCCATGAGAGTTCGGGTTGGACACTGGCCCTG 403
 DB 101 ProGluProValAspLysValAlaValArgLysPheArgValLeuHisThrAlaLeu 120
 QY 404 CACAGAGCTCTCTTACAGGAGCGGCTTTTAAAGTGTCTAGCAATAAGAGTCTCTG 463
 DB 121 HisSerSerSerTyrArgGluAlaValPheLysMetLeuSerAspLysGluSerLeu 140
 QY 464 GATCAGATCATTTGGCCACCCAGGCTCAGAGTACCCCTATGCTCTTGGGTTCTC 523
 DB 141 AspGlnIleValAlaThrProGlyLeuSerSerAspProIleAlaLeuGlyValLeu 160
 QY 524 CAGGACAGGACCTCTCTCTGCTTCTGCTCATCCCAATATGCTTGATAGCTGGTGCCT 583

DB 161 GlnAspLysAspLeuPheSerValPheAlaAspProAsnMetLeuAspThrLeuValPro 180
 QY 584 GCTCACCCAGCCCTCGTCAATGCCATTCCTGTTCTGCACTCCGTAGCAGGAGTCCC 643
 DB 181 AlaHisProAlaLeuValAsnAlaIleValLeuValLeuHisSerValAlaGlySerAla 200
 QY 644 CCAATGCTGGAGTGAATCTCTTCCCGAGAGTCCCTCAGCTCATACGGGATATG 703
 DB 201 ProMetProGlyThrAspSerSerArgSerMetProSerSerSerTyrArgAspMet 220
 QY 704 CCAGGTGGCTTCTCTTGAAGGGCTCTCAGATGATGAGGATGCTTTCACCCAAACACC 763
 DB 221 ProGlyGlyPheLeuPheGluGlyLeuSerAspAspGluAspPheHisProAsnThr 240
 QY 764 AGGTCCACACCTCTAGCAGTACTCCAGCTCCCGCCAGCCCTCCCTGGGTACAGTGA 823
 DB 241 ArgSerThrProSerSerSerSerThrProSerArgProAlaSerLeuGlyTyrSerGly 260
 QY 824 GTGTGTGGCCCGCCGCCATCACCCAGAGTACGTCGCCACCGCTTGGCCCTGGCCAGC 883
 DB 261 AlaAlaGlyProArgProIleThrGlnSerGluLeuAlaThrAlaLeuAlaLeuAlaSer 280
 QY 884 ACTCCGGAGAGCAGCTCTCACACACCGACTCTCTGGCACCCAGGGTCACTTCTCAGGGACC 943
 DB 281 ThrProGluSerSerSerSerHisThrProThrProGlyThrGlnGlyHisSerSerGlyThr 300
 QY 944 TCACCAATGCTCTCTGTTGTCAGTACGGAGCGCCATCACCAATGATCTCTTCCAGCCAA 1003
 DB 301 SerProMetSerSerGlyValGlnSerGlyThrProIleThrAsnAspLeuPheSerGln 320
 QY 1004 GCCCTACAGCATGCCCTTCAGGCTCTGGGAGCCAGCCAGCTTCAGAGCAGTGGCAGGCC 1063
 DB 321 AlaLeuGlnHisAlaLeuGlnAlaSerGlyGlnProSerLeuGlnSerGlnTTPGlnPro 340
 QY 1064 CAGCTGCCAGCAGTACGTACATGGGCATCCAGGACGATGAGCTCAGCTCGGCCCTGC 1123
 DB 341 GlnLeuGlnGlnLeuArgAspMetGlyIleGlnAspAspGluLeuSerLeuArgProCys 360
 QY 1124 AGCCACCGGTGGGACATCCAGCAGCCCTGGAGCTCATCTTTCTGGAGAGCCCAT 1183
 DB 361 ArgProProValGlyThrSerLysGlnProTTPSerSerSerLeuGluGluProHis 380
 QY 1184 GAATCCCTGCTTCCCTGACCCCGCCAGCCAGCTTCAGAGGCTACTGCGCTTGGAGGCA 1243
 DB 381 GluLeuProAlaSerProGluProProAlaSerCysArgGlyTyrCysProTTPGluAla 400
 QY 1244 CTCATGAAGTGGCTCCCATCTCTCCCT 1270
 DB 401 LeuMetLysValProProSerLeuPro 409

RESULT 3

AAV42382

ID AAY42382 standard; Protein; 420 AA.

XX AC AAY42382;

XX DT 09-DEC-1999 (first entry)

XX DE Amino acid sequence of fx317_11.

XX secreted protein; cDNA library; clone; transmembrane protein;
 signal sequence cloning; hybridization cloning; gene therapy;
 receptor.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Peptide 229..241

XX FT Protein /label= Leader/Signal peptide

XX FT Protein 242..420

XX FT Protein /label= Mature protein

XX PN W09942470-A1.

vasotropic; antiparkinsonian; immunostimulant; dermatological;
antithematic; antitumor; antidiabetic; osteopathic; tranquilizer;
cerebroprotective; cytokine; cell proliferation; cell differentiation;
immune deficiency; severe combined immunodeficiency; SCID; tumor;
autoimmune disorder; multiple sclerosis; rheumatoid arthritis;
graft-versus-host disease; myeloid deficiency; wound healing; ulcer;
periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;
Parkinson's disease; Huntington's disease; infection; cardiac disease;
stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;
food supplement; vaccine.

Homo sapiens.

WO2001:75068-A2.

11-OCT-2001.

22-MAR-2001; 2001WO-US09369

30-MAR-2000; 2000US-059330

04-DEC-2000; 2000US-0729674.

(GENY) GENETICS INST INC.

Jacobs K, McCoy JM, Lavallie E, Collins racie LA, Evans C,

Treacy M, Agostino MJ, Steininger RC, Spaulding V, Wong GO,

Clark H, Fichtel K, Verberg G;

WPI; 2001-639363/73.

N-PSDB; AAS5922.

Secreted human proteins, useful as vaccine for treating various
diseases such as autoimmune disorders (e.g. multiple sclerosis), and
nervous system disorders (e.g. stroke).

Disclosure: Page 578-579; 619pp; English.

The invention relates to novel human secreted proteins, the nucleic
acids encoding them. The protein may exhibit cytokine, cell proliferation
or cell differentiation activity or may induce production of other
cytokines in certain cell populations and may exhibit immune stimulating
or immune suppressing activity, which is useful for the treatment of
various immune deficiencies and disorders e.g. severe combined
immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,
systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
inflammation. The proteins are also useful in the treatment of diseases
and disorders including tissue, skin and bone regeneration and in
graft-versus-host diseases (GVHD) in the induction of tumor immunity,
myeloid or lymphoid cell deficiencies, wound healing and tissue repair,
in the treatment of burns, infections and cancers, as well as in treatment
of periodontal disease, osteoporosis or osteoarthritis, mediated by
inflammatory processes, diseases of the peripheral nervous system,
Alzheimer's, Parkinson's disease, Huntington's disease,
amyotrophic lateral sclerosis, and Shy-Drager syndrome, infections,
infarction of cardiac and central nervous system vessels, e.g. stroke,
sepsis, inflammatory bowel disease, ulcers, bone regeneration. The
protein, having activin- or inhibin-related activities is useful as a
contraceptive based on the ability of inhibins to decrease fertility in
female mammals and decrease spermatogenesis in male mammals. The
proteins and nucleic acids are also useful as food supplements. The
present sequence represents a secreted protein of the invention.

SQ Sequence 420 AA;

Alignment Scores:

Pred. No.:	8,71e-162	Length:	420
Score:	1959.00	Matches:	388
Percent Similarity:	99.74%	Conservative:	0
Best Local Similarity:	99.74%	Mismatches:	1
Query Match:	85.10%	Indels:	1
DB:	22	Gaps:	0

US 09-735-251-3 (1-1273) x AA039074 (1-420)

QY	17	AGAAAGAGAGAGCCGAAAGAGAGAGAGATCTCTCTCACACTGGCACCTGGCGGTGAAC	76
DB	32	Arglysa:qGluatGg.uA:Gg.uA:GMetSerLeuSerAspTrpHisLeuAlaValLys	51
QY	77	CTGGCTGACAGGCACCTACTCCAAACTCTATCTTCGGTTGGCAGAGACAGACATGGGA	136
DB	52	LeuAlaAspG.nPrLeuThrProLysSerileLeuArgLeuProGluThrGluLeuGly	71
QY	137	GAATACTGCTAGCGGCTATAGTATTCATTCTGAAGCAGCTTATTGTCGCAACTC	196
DB	72	GluTyrSerLeuGlyG.yTyrSerLeSerPheLeuLysGlnLeuLeAlaGlyLysLeu	91
QY	197	CAGAGTCTGTTCCAGACCTGAGCTGATTCATCTGATCTACTGTCGTCGAGAGCTAAAA	256
DB	92	GlnGluSerValProAspProGluLeuLeAspLeuileTyrGlyArgLysLeuLys	111
QY	257	CATCACCAGACACTGACTTCTATGGCATTCACCTGGTCCACTGTCCTGATGTCGGA	316
DB	112	AspAspGlnThrLeuAspPheTyrGlyileGlnProGlySerThrValHisValLeuArg	131
QY	317	AAGTCTGGCCTGAACCTGATCAGAAACCCGAACTCTGCAAAAGTGGCTGCATGAGA	376
DB	132	LysSerTrpProGluProAspGlnLysProGluProValAspLysValAlaAlaMetArg	151
QY	377	CAGTTCGGGTGTTGCACACTGCCCTGCACAGAGCTCTCTTACAGGAGGCGGTCTTT	436
DB	152	GluPheArgValLeuHisThrAlaLeuHisSerSerSerSerTyrArgGluAlaValPhe	171
QY	437	AAGATGTCACGCAATAGGAGTCTCTCGATCAGATCATTTGGCCACCCAGGCTCAGC	496
DB	172	LysMetLeuSerAsnLysGluSerLeuAspGlnlelleValAlaThrProGlyLeuSer	191
QY	497	AGTCACCTTATGCTCTTGGGGTTCTCCAGGACAAGGACCTCTTCTCTGCTGCTGAT	556
DB	192	SerAspProleAlaLeuGlyValLeuGluAspLysAspLeuPheSerValPheAlaAsp	211
QY	557	CCCATA:GCTTGATAGTGTGGTGGCTGCTCACCAGCCCTCGTCAATGCATGCTGCTG	616
DB	212	ProAsrMetLeuAspThrLeuValProAlaHisProAlaLeuValAsnAlaileValLeu	231
QY	617	GTTCTGCACTCCGTAGCAGGAGTGGCCCCAATGCTGGGACTGACTCTCTTCTCCCGAGC	676
DB	232	ValLeuHisSerValAlaGlySerAlaProMetProGlyThrAspSerSerArgSer	251
QY	677	ATGCCCTGACCTCATACCGGATATCCAGGTGGCTCTCTGTTTGAAGGGCTCTCAGAT	736
DB	252	MetProSerSerSerTyrArgAspMetProGlyGlyPheLeuPheGluGlyLeuSerAsp	271
QY	737	GATGAGGATGCTTTCACCCCAACACAGGTCACACCCCTTAGCAGTATCCGAGCTCC	796
DB	272	AspGluAspAspPheHisProAsnThrArgSerThrProSerSerSerThrProSerSer	291
QY	797	CGCCAGGCTCTCCCTGGGTACAGTGGAGTCTGGGCCCCGGCCCATCACCAGAGTGAG	856
DB	292	ArgProAlaSerLeuGlyTyrSerGlyValAlaGlyProArgProleThrGlnSerGlu	311
QY	857	CTGGCACCCTGGCTGGCCCTGGCCAGACCTCCGAGAGAGAGCTCTCACACCGACTCTCT	916
DB	312	LeuAlaThrAlaLeuAlaLeuAlaSerThrProGluSerSerSerHisThrProThrPro	331
QY	917	GGCACCAGGCTCATCTCAGGAGCCTCACCATGCTCTGCTGTCAGTCCAGGAGCAG	976
DB	332	GlyThrGlnGlyHisSerSerGlyThrSerProMetSerSerGlyValGlnSerGlyThr	351
QY	977	CCCATCACAATGATCTCTTTCAGGCAAGCCCTACAGCATGCCCTTACGGCTCTGGGCGAG	1036
DB	352	ProleThrAsnAspLeuPheSerGlnAlaLeuGlnHisAlaLeuGlnAlaSerGlyGln	371
QY	1037	CCAGCCTTCAGAGCCAGTGGCCGCCCGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	1096
DB	372	ProSerLeuGlnSerGlnTrpGlnProGlnLeuGlnGlnLeuArgAspMetGlyileGln	391

QY 497 AGTGACCTATTGCTTGGGGTCTTCAGGACAGGACCTCTTCTCTGCTTGGCTGAT 556
 DB 192 SerAspProIleAlaLeuGlyValLeuGlnAspYsAspLeuPheSerValPheAlaAsp 211
 QY 557 CCCATATGCTTGATACCTGGTGGCTGCTACCCAGGCTCTGCTCAATGCGATGCTG 616
 DB 212 ProAsnMetLeuAspThrLeuValProAlaHisProAlaLeuValAsnAlaLeValLeu 231
 QY 617 GTTCTGCACTCCGTAGCAGGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 676
 DB 232 ValLeuHisSerValAlaGlySerAlaProMetProGlyThrAspSerSerSerArgSer 251
 QY 677 ATGCCCTCAGCTCATACCGGATATGCCAGTGGCTTCCGTTTGAAGGCTCTCAGAT 736
 DB 252 MetProSerSerSerTyrArgAspMetProGlyGlyPheLeuPheGlyGlyLeuSerAsp 271
 QY 737 GATGAGGATGCTTTCACCAACACACAGGTCACACCTCTACGACTACTCCAGCTCC 796
 DB 272 AspGluAspAspPheHisProAsnThrArgSerThrProSerSerSerThrProSerSer 291
 QY 797 CGCCAGCTCCCTGGGTACAGTGGAGCTCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 856
 DB 292 ArgProAlaSerLeuGlyTyrSerGlyAlaAlaGlyProArgProIleThrGlnSerGln 311
 QY 857 CTGGCCAGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 916
 DB 312 LeuAlaThrAlaLeuAlaLeuAlaSerThrProGluSerSerSerHisThrProThrPro 331
 QY 917 GGCACCCAGGCTCATCTCTCAGGACCTCCACCAATGCTCTGCTGCTGCTGCTGCTGCT 976
 DB 332 GlyThrGlnGlyHisSerSerGlyThrSerProMetSerSerGlyValGlnSerGlyThr 351
 QY 977 CCATCAACCAATGATCTCTTCAGGCAAGCCCTATAGCATGCCCTTCAGGCTCTGGGCGAG 1036
 DB 352 ProIleThrAsnAspLeuPheSerGlnAlaLeuGlnHisAlaLeuGlnAlaSerGlyGln 371
 QY 1037 CCCAGCTTCAGAGCCAGTGGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1096
 DB 372 ProSerLeuGlnSerGlnThrProGlnLeuGlnLeuAlaGlnMetGlyIleGln 391
 QY 1097 GACGATGAGTGGAGCTCCGGCTCAGGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1155
 DB 392 AspAspGluLeuSerLeuArgAlaLeuGlnAlaThrGlyGlyAspIleGlnAlaAlaLeu 411
 QY 1156 GAGCTCATCTTGTGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1214
 DB 412 GluLeuIlePheAlaGlyAlaAlaPro 426

RESULT 6
 AAG93292
 ID AAG93292 standard; Protein: 380 AA.
 XX AC AAG93292;
 AC AC AAG93292;
 DT 13-SEP-2001 (first entry)
 XX DE Human protein HP:0657.
 XX KW Human; gene therapy; tumour.
 XX OS Homo sapiens.
 XX PN WO2001:42302-A1.
 XX PD 14-JUN-2001.
 XX PF 06-DEC-2000; 2000WO-JP08631.
 XX PR 06-DEC-1999; 930JP-0346863.
 XX PR 06-DEC-1999; 95JP-0346864.
 XX PR 08-FEB-2000; 2000JP-0031062.
 XX PR 10-FEB-2000; 2000JP-0034090.

PR 10-FEB-2000; 2000JP-0034091.
 PR 14-FEB-2000; 2000JP-0035829.
 PR 14-FEB-2000; 2000JP-0035899.
 PR 14-MAR-2000; 2000JP-0071161.
 PR 30-MAY-2000; 2000JP-0160851.
 XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
 PA Kato S, Eguchi C, Saeki M;
 XX WPI: 2001-381646/40.
 DR N-PSDB; AAH68577.
 XX Human protein originated from tumor cell line, applicable as drug,
 PT reagent for studying intracellular protein networks and protein source
 PT for drug screening, also encoded cDNA for gene diagnosis and gene
 PT therapy
 XX Claim 1: Pages 313-314; 471pp; Japanese.
 CC The present sequence is a human protein. The human protein, preferably
 CC originated from tumor cell line, is applicable as a drug, a reagent for
 CC studying intracellular protein networks and a protein source for
 CC screening proteins for binding low molecular weight drugs. The human
 CC protein coding sequence is useful for gene diagnosis and gene therapy,
 CC expression vectors and transformant cells for detection of ligands and
 CC receptors.
 XX SQ Sequence 380 AA;
 Alignment Scores:
 Pred. No.: 2,31e-157 Length: 380
 Score: 1908.00 Matches: 378
 Percent Similarity: 99.47% Conservative: 0
 Best Local Similarity: 99.47% Mismatches: 2
 Query Match: 82.88% Indels: 1
 DB: 22 Gaps: 0
 US-09-735-251-3 (1-1273) x AAG93292 (1-380)
 QY 44 ATGTCTCTCTCAGACTGGCACCCTGGCGGTGAAGCTGGCTGACCCAGCCACTTACTCCAAG 103
 DB 1 MetSerLeuSerAspTrpHisLeuAlaValLysAlaAspGlnProLeuThrProLys 20
 QY 104 TCTATTCTTCGGTTCGCCAGACAGAACTGGGAGAACTACTCCTAGGGGGCTATAGTATT 163
 DB 2 SerIleLeuArgLeuProGluThrGluLeuGlyGlySerLeuGlyGlyTyrSerIle 40
 QY 164 TCATTTCGAAACAGCTTATTGCTGGCAAACTCCAGGAGTCTGTTCACAGACCTGAGCTG 223
 DB 41 SerPheLeuGlySGIleuIleAlaGlyLysLeuGlnGluSerValProAspProGluLeu 60
 QY 224 ATTGATCTGATCTACTGTGTCGGAAGCTAAAGATGACACAGACACTTGTCTTATCGC 283
 DB 61 IleAspLeuIleTyrCysGlyArgLysLeuLysAspAspGlnThrLeuAspPheTyrGly 80
 QY 284 ATTCACCTGGTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 343
 DB 81 IleGlnProGlySerThrValHisValLeuArgLysSerTrpProGluProAspGlnLys 100
 QY 344 CCGGAACCTGTGGCAAAAGTGGCTGCCATGAGAGAGTTCGCGGTGCTTCCACACTGCCCTG 403
 DB 101 ProGluProValAspLysValAlaAlaMetArgGluPheArgValLeuHisThrAlaLeu 120
 QY 404 CACAGAGCTCCTTTACAGGAGCGCTCTTTAAGATGCTCAGCAATTAAGAGTCTCTG 463
 DB 121 HisSerSerSerTyrArgGluAlaValPheLysMetLeuSerAsnLysGluSerLeu 140
 QY 464 GATCAGATCATCTGTCGCCACCCAGGCTTCAGAGTGCCTTATGCTCTTGGGGTCTC 523
 DB 141 AspGlnIleIleValAlaThrProGlyLeuSerSerAspProIleAlaLeuGlyValLeu 160
 QY 524 CAGGACAGGAGCTCTCTCTGCTCTCTGCTGATCCCAATAAGCTTGATACCTGCTGCT 583

DR WP1: 2001-025274/03.
XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
XX cancer.
XX
XX Example 1; Page 790; 799pp; English.
XX
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAPs) respectively.
CC AAF23015 to AAF23018, AAF23019 to AAF23022 and AAF23023 to AAF23026
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytosolic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterized by expression of an abnormal amount of a protein,
XX e.g. cancer.
XX
XX Sequence :110 AA;
SQ
Alignment Scores:
Pred. No.: 1,59e-33 Length: 110
Score: 483.00 Matches: 93
Percent Similarity: 88.99% Conservative: 4
Best Local Similarity: 85.32% Mismatches: 11
Query Match: 20.49% Indels: 2
DB: 22 Gaps: 0
US-09-735-251-3 (1:1273) x AAF23051 (1:110)
QY 934 GAATGACCTCGGTCGACGAGTGGTGTGACAGTGTCTCCGAGTGTGGCCAG 875
Db 1 GlyMetThrLeuGlyAlaArgSerArgCysValArgAlaLeuArgSerAlaGlyLeu 20
QY 974 GCGCAAGCGGGTGGCAGTCACTTGGGTGATGGCGAGTCCGAGGCTCCACTGA 815
Db 21 GlyGinglyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 40
QY 814 CCCCAGGAGGTGGCGGAGTATGGTATGAGTATGAGTATGAGTATGAGTATGAGT 756
Db 41 ProGnglyGlyTrpAlaGlyAlaGlySerThrAlaHisAlaValTrpThrCysLeu 60
QY 755 GGTGAAGTCACTCCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 696
Db 61 GlyGlySerHisProHisHisHisHisHisHisHisHisHisHisHisHisHis 80
QY 695 GGTATGAGTGGAGGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTAT 636
Db 80 oValLeuAlaGlyHisAlaHisAlaHisAlaHisAlaHisAlaHisAlaHisAla 100
QY 635 CTGCTACGAGTGCAGAACCGAGAC 611
Db 100 acysPheArgValAsnGlyGlu 108
RESULT 13
AAB58780
ID AAB58780 standard; Protein: 114 AA.
XX
XX AAB58780;
XX
XX 27-MAR-2001 (first entry)
XX
XX Breast and ovarian cancer associated antigen protein sequence SEQ ID 488.
XX
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neutrotic; neuroprotective; antiviral; antiparasitic and cardiotropic;
KW antidiabetic; antiinflammatory; antitumor; vulnerable; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW

KW cardiovascular disorder; wound healing; neurological disease.
XX
XX Homo sapiens.
XX
XX WO200055173-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WC-US05581.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-611515/58.
XX
XX N-PSDB; AAF21683.
XX
XX
XX New human breast and ovarian cancer associated gene sequences and the
XX polypeptides encoded by these genes, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases.
XX
XX Claim 11; Page 919; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX associated with breast and ovarian cancer. Included in the invention are
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX isolation and characterization of the DNA and protein sequences of the
XX invention. The breast and ovarian cancer associated DNA, protein, agonist
XX or antagonist sequences exhibit cytostatic; immunosuppressive;
XX neutrotic; neuroprotective; antiviral; antiparasitic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; vulnerable; anticonvulsant;
XX antibacterial; antifungal; antiparasitic and cardiotropic activity. The
XX polynucleotide and protein sequences are used in the diagnosis of cancer,
XX particularly breast and ovarian cancer. The nucleic acid sequences,
XX proteins, agonists and antagonists may also be used in the diagnosis,
XX prevention and treatment of immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; cardiovascular disorders such as
XX myocardial ischaemia; wound healing; neurological diseases such as
XX cerebral anoxia and epilepsy; and infectious diseases.
XX
XX Sequence 114 AA;
SQ

Alignment Scores:
Pred. No.: 3,99e-30 Length: 114
Score: 444.00 Matches: 92
Percent Similarity: 88.79% Conservative: 3
Best Local Similarity: 85.98% Mismatches: 10
Query Match: 19.29% Indels: 2
DB: 21 Gaps: 0

US-09-735-251-3 (1:1273) x AAB58780 (1:114)

QY 8 GGCAGAGGAG 67
Db 10 GlyThrGlyArgGlyArgGlyArgGlyArgGlyArgGlyArgGlyArgGlyArg 29
QY 68 CCGGTGAAGCTGGCTGACAGGACCTTACTCCAAAGTCTATCTTCGTTGCCAGACA 127
Db 30 AlaValLysLeuAlaAspGlnProLeuAlaProLysSerLeuGlnLeuProGluSer 49
QY 128 GAATGGAGAGATCTCGTAGGGGCTATAGTATTATTCATTCGAGCAGCTTATGCT 187
Db 50 GluLeuGlyGlyTrpSerLeuGlyGlyTrpSerLeuGlyGlyGlyGlyGlyGlyGly 69
QY 188 GGCAGAACCTCCAGGAGTCTGTTCCAGACCCCTGAGCTGCTATTCATCTGCTGCTG 247
Db 70 GlyLysLeuGlnGluSerValProAspProGluLeuLeuLeuLeuLeuLeuLeuLeu 99

QY 248 AAGCTAAAGATGACGACAGACTTCATCGCATTCAGCTGGCTGCCTGCTCAT 307
 Db 90 LysLeuLysAspAsp...Thr-LeuThrSerThrVal-PheAsnLeuAlaPro-HisProC 109
 QY 308 GTTCTGCGAAAGCTCCTG 324
 Db 109 ysSer...Glu...Leu 114

RESULT 14

AAG02759
 ID AAG02759 standard; Protein: 76 AA.

XX AAG02759;
 AC

DT 06-OCT-2000 (first entry)
 XX

DE Human secreted protein, SEQ ID NO: 6840.
 XX

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX

OS Homo sapiens.
 XX

PN EP1033401-A2.
 XX

PD 06-SEP-2000.
 XX

PF 21-FEB-2000; 2000EP-0203610.
 XX

PR 26-FEB-1999; 99US-0122487.
 XX

PA (GEST) GENSET.
 XX

PI Dumas Milne Edwards J, Ducloux A, Giordano J;
 XX

DR WPI; 2000-500381/45.
 DR

DR N-PSDB; AAC02765.
 XX

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PT Claim 13; SEQ ID 6840; 71pp + CD-ROM; English.
 PS

XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or poly(A) RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (3'UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC 3'UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX

SQ Sequence 76 AA;

Alignment Scores:

Pred. No.:	Length:	T _n
Score:	2,098-25	76
Percent Similarity:	389.00	76
Best Local Similarity:	100.00%	Conservative:
Query Match:	100.00%	Mismatches:
DB:	16.90%	Indels:
	21	Gaps:
		0

US-09-735-251-3 (1-1273) x AAG02759 (1-76)

QY 44 ATGTCTCTCTCAGCTGGCAGCTTCGCTGAGCTGGCTGACCGCAGCTTACTCCAAAG 103
 |||||

Db 1 MetSerLeuSerAspTrpHisLeuAlaValLysLeuAlaLeuGlnProLeuThrProLys 20
 QY 104 TCTATCTTCGTTCCGACAGACAGACTGGGAGAACTCTCGTAGCGGCTATAGTATT 163
 Db 21 SerIleLeuArgLeuProGluThrGluLeuGlyGlyTyrSerLeuGlyTyrSerIle 40
 QY 164 TCATTTCGAAAGCAGCTTATTGCTGGCAAACTCCAGGAGTCTGTTCCAGACCTCGAGCTG 223
 Db 41 SerPheLeuLysGlnLeuIleAlaGlyLysLeuGlnGluSerValProAspProGluLeu 60
 QY 224 ATTGATCTGATCTACTGTGTCGGAAGCTAAAGATGACGACAGACTT 271
 Db 61 IleAspLeuIleTyrCysGlyArgLysLeuLysAspAspGlnThrLeu 76

RESULT 15

ABB69125

ID ABB69125 standard; Protein: 547 AA.

XX ABB69125;
 AC

DT 26-MAR-2002 (first entry)
 XX

DE Drosophila melanogaster polypeptide SEQ ID NO 34167.
 XX

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX

OS Drosophila melanogaster.
 XX

PN WO200171042-A2.
 XX

PD 27-SEP-2001.
 XX

PF 21-MAR-2001; 2001WC-US09231.
 XX

PR 23-MAR-2000; 2000US-191637P.
 PR

PR 11-JUL-2000; 2000US-0614150.
 XX

PA (PEKE) PE CORP NY.
 XX

PI Venter JC, Adams M, Li PWD, Myers EW;
 XX

DR WPI; 2001-656860/75.
 DR

DR N-PSDB; ABU13228.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PS Disclosure; SEQ ID NO 34167; 21pp + Sequence Listing; English.
 XX

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 CC sequences (ABU1840-ABU16175) and the encoded proteins
 CC (ABU57737-ABU72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 547 AA;

Alignment Scores:

Pred. No.:	Length:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:	
3,16e-06	547	172.50	35.30%	20.15%	7.49%	22	
Matches:	109	Conservative:	82	Mismatches:	153	Indels:	198
Gaps:	23						

US-09-735-251-3 (1-1273) x ABB69125 (3 547)

```

QY 95 ACTCCAAAGTCTATTCTTCGGTTCCACAGACAGACTGGGAGATATCTCGTAGGGGC 154
Db |||||
QY 15 ThrProLysAsp-----LysLysThrValGluValAspGluAspSer 28
Db |||||
QY 155 TATATATTTCATTCTTAAGCACTTATTGTGCAAACTCCAGAGAGTGTTCACAGAC 214
Db |||||
QY 29 ---GlyIleLysAspPheLysIleLeuValAlaGlnLysProGlu 44
Db |||||
QY 215 CTGAGCTGATTGATCTGATCTA TGTGCTGGAGAGCTAAAGATGACAGACACTTGAC 274
Db |||||
QY 45 ProGluGlnLeuValLeuIlePheAlaGlyLysIleMetLysAspThrAspThrLeuGln 64
Db |||||
QY 275 TTCTATGGCATTCAACCTGGTCCACTGTCCTGCTGCAAGTCTGCAAGTCTGCTGAACCT 334
Db |||||
QY 65 MetHisAsnIleLysAspAsnLeuThrValHisLeuValIleLysAlaProThrArgAsn 84
Db |||||
QY 335 GATCAGAAACCG-----CAACCTGTGGAC----- 358
Db |||||
QY 85 AsnGluGlnProAlaArgGlnProAlaAspValArgGlnThrProPheLysLeuAsnGln 104
Db |||||
QY 359 -----AAGTGGCTGCCATGAGA----- 376
Db |||||
QY 105 PheGlyGlyLeuAlaGlyMetGluAlaLeuGlyAlaGlySerAsnThrPheMetAspLeu 124
Db |||||
QY 377 GAGTTCGGGGTGTTCACACTGCTGTCACACACAGCTCTCTCTACAGGAG----- 427
Db |||||
QY 125 GlnAlaArgMetGlnAsnGlnLeuLeuAsnAsnGlyAspValLeuArgSerLeuMetAsp 144
Db |||||
QY 428 -----GGGGTCTTTAAGATGCTAGCAATAAGGAGTCTCTGATTCAGATCATTTGTGCC 481
Db |||||
QY 145 AsnProMetValGlnGlnMetMetAsnAsnProAspThrValArgGlnLeuIleThrSer 164
Db |||||
QY 482 ACCGACAGCC-----CTAGCAGCTGAGTATGTTCTGGGGTCTCTCCAGGAC 529
Db |||||
QY 165 AsnProGlnMetHisAspLeuMetGlnArgAsnProGluLysSerHisMetLeuAsnAsn 184
Db |||||
QY 530 AAGGACCTC-----TTCTGTCTTGGCTGATGCTCAATGCTGATACGTGTG 577
Db |||||
QY 185 ProAspLeuLeuArgGlnThrMetGluLeuAlaArgAsnProSerValLeuGlnGluLeu 204
Db |||||
QY 578 GTGCTGTCTAC-----CCAGCCCTCGTCAATGCC 607
Db |||||
QY 205 MetArgSerHisAspArgAlaMetSerAsnGlnSerValProGlyLysSerAla 224
Db |||||
QY 608 ATTGCTCTGCTCTGCAATGCT----- 631
Db |||||
QY 225 LeuGlnArgIleThrArgAspProGlnGlnValAlaAlaAlaAlaAlaAlaAlaAla 244
Db |||||
QY 632 -----GCATGAGTCCGCA 646
Db |||||
QY 245 GlyArgAsnProPheAlaGlyLeuValAspGlyGlyGlyValAlaGlyAsnAsnPro 264
Db |||||
QY 647 ATGCTGTGGAGTACTCTCTCTCCGGGAGATGCCC-----TCCAGC 688
Db |||||
QY 265 GlnGlnGlyThrGluAsnArgAsn---ProLeuProAsnProIlePheGlyAlaAsnSer 283
Db |||||
QY 689 TCATACCGGATATGCCAGTGGTCTCTGTTGAGGGCTTCAGATGATGAGGATGAC 748
Db |||||
QY 284 GlyThrAsnGlyThrValGlyGly-----SerGlyValArgLysAsnProThrGlyAsp 300
Db |||||
QY 749 TTTCACCCA---AACACAGGTCCACACCTCT----- 778
Db |||||
QY 301 LeuProProAsnAsnValLeuAsnThrProAlaMetArgSerLeuGlnMetAla 320
Db |||||
QY 778 ----- 778
Db |||||
QY 321 AspAsnProAlaMetMetGlnAsnLeuLeuAsnAlaProTyrThrArgSerMetMetGlu 340
Db |||||
QY 779 -----AGCAGTACTCCC-----AGC 793
Db |||||
QY 341 SerMetSerGlnAspProAspMetAlaAlaArgLeuLeuSerSerProLeuMetSer 360
Db |||||

```

```

QY 794 TCCGCCACAGCTCCCTCGGGGTACAGTGGAGCTGCTGGGCCCGGCCCATCCACAGAGT 853
Db |||||
QY 361 AsnAsnProAlaLeuGlnGluGlnValArgGlnMetMetProGlnPheMetAlaGlnMet 380
Db |||||
QY 854 GAGTGGCCACCCCTTGGCCCTGGCCAGCAGCTCCGGAGAGCAGCTCT----- 901
Db |||||
QY 381 GlnAsnProGluValMetAsnMetLeuThrAsnProAspAlaMetAsnAlaIleLeuGln 400
Db |||||
QY 902 -----CACACACCCACTCTCTGGC----- 919
Db |||||
QY 401 IleGlnGlnGlyMetGluGlnLeuArgSerAlaAlaProGlyLeuValGlyThrLeuGly 420
Db |||||
QY 920 -----ACCCAGGCTCATCTCTCAGGACCTCCACCAATG----- 952
Db |||||
QY 421 IleProProProProGlyAlaGlyThrGlyThrAsnProAlaSerGlyAspGlySer 440
Db |||||
QY 953 -----TCTCTGTGTCTCCAGTCAGG 973
Db |||||
QY 441 GlyGlyAsnSerGlyAlaSerThrAsnAsnValSerProSerSerGlyLeuAsnAlaGly 460
Db |||||
QY 974 ACGCCCATCACCAT-----CATCTCTCAGCCAGGCC 1006
Db |||||
QY 461 ThrGlyThrProAsnLeuAlaProGlyGlyGlyProAsnAlaGlnLeuPheAsnAspPhe 480
Db |||||
QY 1007 CTACAGCATGCTCTTCAGGCC-----TCTGGCAGCCCGCCAGCTTCAG 1048
Db |||||
QY 481 MetMetArgMetLeuAsnGlyMetSerAsnAsnAlaAspAsnThrGlnProPro---Glu 499
Db |||||
QY 1049 AGCAGTGGCAGCCCGCAGCTGCGAGCAGCTACGTGATCGGATCCAGCAGATGAGCTG 1108
Db |||||
QY 500 ValArgTyrGlnSerGlnLeuGluGlnLeuAsnAlaMetGlyPheValAsnArgAspAla 519
Db |||||
QY 1109 AGCTGCG-CCCTGCGAGCCCGGTCGGGAGCATCCAGCAGCCCTGGAGCTCATCTTT 1167
Db |||||
QY 520 AsnLeuGlnAlaLeuIleAlaThrPheGlyAspIleAsnAlaAlaValGluArgLeuLeu 539
Db |||||
QY 1168 GCT 1170
Db |||||
QY 540 Ser 540

```

Search completed: November 12, 2003, 06:46:17
Job time : 55 secs

112-00-325-2 (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838,

JS-09-735-251-3 (1-1273) x US-09-890-688-86 (1-380)

QY 44 ATGTCCTCTCAGACTGCGACCTGCGGTGAAGTGGCTGACAGCCACTTACTCCAAAG 163
DB 1 MetSerLeuSerAspTyrHisLeuAlaValLysLeuAlaAspGlnProLeuThrProLys 20
QY 104 TCTATCTTCGGTCCGACAGACAACTGGAGAAATACGGTAAAGGGGGCTATAGTATT 163
DB 21 SerIleLeuArgLeuProGluThrGluLeuGlySerLeuGlySerLeuGlySerIle 40
QY 164 TCATTTCTGAAGACCTTATCTGCTGAACTGAGGAGTCTCTCCAGACCTGACCTG 223
DB 41 SerPheLeuLysGlnLeuIleAlaGlyLysLeuGlnGlySerValProAspProGluLeu 60
QY 224 ATTGATCTGATCTACTGCTGGAGAGTAAAGATGACAGACCTTACTTCTATGGC 283
DB 61 IleAspLeuIleTyCysGlyArgLysLeuLysAspGlnThrLeuAspPheTyGly 80
QY 284 ATTCAAGCTGGTCCACTGTCATCTCTGCGAAGTCTGGCTGACCTGATCATCAAA 343
DB 81 IleGlnProGlySerThrValHisValLeuArgLysSerTyrProGluProAspGlnLys 120
QY 344 CGGAACCTGTGGCAAAAGTCTGCTGATGAGAGTCTGGGTGTTCCACACTGGCCTG 403
DB 121 ProGluProValAspLysValAlaAlaValArgGlnPheArgValLeuHisThrAlaLeu 120
QY 404 CACAGCAGCTCTCTTACAGGAGAGGCTCTTGAAGTCTGACCAATAGATCTCTG 463
DB 121 HisSerSerSerTyrArgGluAlaValPheLysValLeuSerAsnLysGluSerLeu 140
QY 464 GATCAGATCATTTGGCCACCCCACTGCTGAGTACCTGATTCCTGCTGGGTCTC 523
DB 141 AspGlnIleIleValAlaThrProLysSerSerAspProIleAlaLeuGlyValLeu 160
QY 524 CAGCAAGAGGACCTCTCTCTCTCTGCTGATGCAATATGATGATGATGCTGCT 593
DB 161 GluAspLysAspLeuPheSerValPheAlaAspProAsnValLeuAspThrLeuValPro 180
QY 594 GCTCAGCAGGCTGCTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 643
DB 181 AlaHisProAlaLeuValAspAlaIleValLeuValLeuHisSerValAlaGlySerAla 200
QY 644 CCAATGCTGGGACTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 703
DB 201 ProMetProGlyThrAspSerSerSerSerSerSerSerSerSerSerSerSer 220
QY 704 CAGGTGCT 763
DB 221 ProGlyGlyPheLeuPheLeuGlyLeuAlaValLeuAlaValLeuAlaValLeu 240
QY 764 AGGTCCACAGCTCTACAGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 823
DB 241 ArgSerThrProSerSerSerSerSerSerSerSerSerSerSerSerSerSer 260
QY 824 GCTGCTGGGCCCCCGCCCATCAGCAAGTGAAGTGGCCAGCCGCTTGGCCCTGCGCAGC 883
DB 261 AlaAlaGlyProArgProIleThrGlnSerGlnLeuAlaAlaLeuAlaLeuAlaSer 280
QY 884 ACTCCGAGAGAGGCTCTCAGACAAGACTCTGCGACCCCAAGTGTATCTCTCAGGAC 943
DB 281 ThrProGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 300
QY 944 TCACCAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1003
DB 301 SerProMetSerSerSerGlyValGlnSerGlyThrProIleThrAsnAspLeuPheSerGln 320
QY 1004 GCCCTACAGCATGCCCTTTCAGGCCCTCTGGGAGCCCGCCAGCCCTTACAGCCATG 1063
DB 321 AlaLeuGlnHisAlaLeuGlnAlaSerGlyGlnProSerLeuGlnSerGlnPro 340
QY 1064 CAGCTGCGACGACTACTGATCATGCGGCAATCCAGACAGTATGAGCTGCGGCCCTG 1122
DB 341 GlnLeuGlnGlnLeuArgAspMetGlyIleGlnAspAspGlnLeuSerLeuArgAlaLeu 360

QY 1123 CAGGCACACCGTGGGACATCCAGACGCTCCAGAGCTCATCTTTCTGGAGGACCCCA 1182
DB 361 GlnAlaThrGlyGlyAspIleGlnAlaAlaLeuGluLeuIlePheAlaGlyGlyAlaPro 380
RESULT 3
US-10-102-806-488
; Sequence 488, Application US/10102806
; Publication No. US2003054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P033FIC1
; CURRENT APPLICATION NUMBER: US/10/102,806
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US06/05881
; PRIOR FILING DATE: 2003-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 488
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (95)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (111)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (113)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-488
Alignment Scores:
Pred. No.: 1,096-23 Length: 114
Score: 444.00 Matches: 92
Percent Similarity: 88.75% Conservative: 3
Best Local Similarity: 85.98% Mismatches: 10
Query Match: 19.23% Indels: 2
DB: 15 Gaps: 0
US-09-735-251-3 (1-125): x US-10-102-806-488 (1-114)
QY 8 GCCCAG 67
DB 10 GlyThrGlyArgGlyArgGlyArgGlyArgGlyArgGlyArgGlyArgGlyArg 29
QY 68 GCGGTGAAGCTGGCTGACCCAGCCACTTACTCTCAAGTCTATTTCTCGGTGCCAGAGACA 127
DB 30 AlaValLysLeuAlaAspGlnProLeuAlaProLysSerIleLeuGlnLeuProGluSer 49
QY 128 GAAGTGGAGAGATATCTGCTAGCGGGCTATAGTATTTCTTCTTCAAGCAGCTTATGCT 187
DB 50 GluLeuGlyGlyTyrSerLeuGlyGlyTyrSerIleSerPheLeuLysGlnLeuIleAla 69
QY 188 GCGCAACTCCAGAGTCTGTTCCAGACCTGAGCTGATTGATCTGATCTGCTGCTCGG 247
DB 70 GlyLysLeuGlnGluSerValProAspProGluLeuIleAspLeuIleTyrCysGlyArg 89
QY 248 AGCTTAAAGATGACAGACACTTCACTTCTATGATTCATCACTTCAAGCTGCTGCTCAT 307
DB 90 LysLeuLysAspAep***ThrLeuThrSerThrValPheAsnLeuAlaPro-HisProC 109
QY 308 GTTCTCGGAAAGTCTCTG 324
DB 109 ysSer***Glu***Leu 114
RESULT 4

```

US-09-745-008-34
; Sequence 34, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09745008
; CURRENT FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: US 60/171,481
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
US-09-745-008-34

Alignment Scores:
Pred. No.: 0.00205 Length: 1162
Score: 163.50 Matches: 109
Percent Similarity: 36.16% Conservative: 43
Best Local Similarity: 24.94% Mismatches: 181
Query Match: 7.10% Indels: 98
DB: 10 Caps: 18

US-09-735-251-3 (1-1273) x US-09-745-008 34 1162:
QY 59 TGGACCTGGCGGTGAAGCTGGTGGACAGCCTTACTCCMAAGTCTATTCTCGGTG 118
DB 499 TyrHisLeuValLeuThrMetAlaAsn 507
QY 119 CCAGACACAGCACTGGGAGAAATATCTGTAGCGGGCTATAGATTTCATTCTGGAGAG 178
DB 508 -----LysIleGlySerValIleAspGly-----Glu 517
QY 179 CTTATTCTCGCAACCTCCAGGAATCTGTGTGACACCTCGAGTATGAT----- 229
DB 518 LeuLeuGluGlySerGlyGlnThrValProAspGlyArgThrProAspIleSerHis 537
QY 229 ----- 229
DB 538 PheTyrValGlyGlyTyrLysAlaSerAspValProThrIleMetThrValAsn 557
QY 230 -----CTGACTTACTTGGTTTAAATTAAGAAATATGACCTTACTCTCTATGSC 283
DB 558 AsnValLeuLeuTyrAsnArgAlaIleAlaValThrIleGluIleAlaGlyThrLeuPheLeu 577
QY 284 ATTCACACCT-----GGTCCACCTGCTCATGTTCTGGGAAAGTCTTGCGCTGAACCTGAT 337
DB 578 SerGlnAspLeuIleGlyThrGlnAlaHisMetAspSerSerSerSerAlaHis 597
QY 338 CAGAAACCGGAA-----CTGTGGAAGAGTGTCCCATGAGAGAGTTCGGGTGTGAC 394
DB 598 SerThrProSerThrProAlaAspSerSerAlaHisSerThrProSerThrProValAsp 617
QY 395 ACTGCTCTGCACACAGCTCTCTATAGGAGAGCGGTCTTAAAGATGCTCAGCAATAAG 454
DB 618 SerSerAlaHisSerThrProSer----- 625
QY 455 GAGTCTCTGGATCAGATCATTTGTGGCACCCCTGAGGCTCAGAGCTGACCTATTGCTTT 514
DB 626 ThrProAlaAspSerSerAlaHisGlyThrProSerThrProValAspSerSerAlaHis 645
QY 515 GGGGTCTCCAGGACAGGAGCTTTCTGTGCTCTGCTGATCCCAATATG-----CTGTAT 571
DB 646 GlyThrProSerThrProAlaAspSerSerAlaHisGlyThrProSerThrProValAsp 665
QY 572 ACGTTGGTGGCTGTCCACCCAGCGCTCTGTCATGCCATGCTGCTGCTGCTGCTGCTGCTG 611
DB ----- 611

```

```

RESULT 5
; Sequence 20, Application US/09843746
; Publication No. US2003016650A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hwei-Mei
; APPLICANT: Honchell, Cynthia D.
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Mucin-Related Tumor Marker
; FILE REFERENCE: PG-0019 US
; CURRENT APPLICATION NUMBER: US/09/840,746
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PERL Program
; SEQ ID NO 20
; LENGTH: 528
; TYPE: PRT
; ORGANISM: Sub scrofa
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Genbank IT No. US2003016650A1 g915208
US-09-840-746-20

Alignment Scores:
Pred. No.: 0.00371 Length: 528
Score: 159.00 Matches: 91
Percent Similarity: 37.02% Conservative: 53

```

[illegible]

```

Db 312 ThrProThrProProLeuArgHisAlaAlaThrArgPheAlaThrSerLeuGlySer 331
QY 398 GCGCTGCAGCAGCTCTCTTACAGGAGGGGCTTTAAGATGCTCAGCAATAAGAG 457
Db 332 AlaPheHisProValLeuProHisThrAlaThrValProArgProLeuAsnLysAsn 350
QY 458 TCTCTGGATCAGATCATTTGGCACCCAGGAGCTTCAGCATGACCTCTCTTGGG 517
Db 351 SerArgProSerProValAsnThrProSerSerGlnProAlaAlaLys 368
QY 518 GTTCTCCAGGACAGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 577
Db 368 SerCysAlaThrProThrSerArgPheSerProLeu 380
QY 578 GTGCTGTCTACCCAGCCCTCTCTAATGCCATTGTCTGTCTGTCTGTCTGTCT 637
Db 381 ProProSerProThrAlaProProSerProThrAlaProProSerProThrAla 395
QY 638 A-----GTGCTCATGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 679
Db 395 aThrGlyProArgProValLeuProValCysValSerSerProValProGlnMe 413
QY 680 CCTCTCAGCTCATACCGGATATGCAATGCTCTCTCTCTCTCTCTCTCTCTCT 739
Db 413 tProProSerProThrAlaProProSerProThrAlaProProSerProThrAla 425
QY 740 GAGGATGACTTTACCCCAACATAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 799
Db 425 pSerValThrThrProValSerProProThrSerGlyProAlaAlaProProPro 445
QY 800 CAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 859
Db 445 oProProProProProProProProProProProProProProProProPro 457
QY 860 GCGAGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 904
Db 457 uProProProProProProProProProProProProProProProProPro 475
QY 905 ACAGGAGCTCTGCGACCGAGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 955
Db 476 SerProProGlyThrProLeuAlaSerProSerSerSerSerSerSerSer 495
QY 956 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1097
Db 496 SerProSerAlaGlyValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1057
QY 998 AGTAAAGCTTACAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1057
Db 516 SerSerAlaSerGlyProGlyLeuAlaAlaAlaAlaAlaAlaAlaAlaAla 533
QY 1058 GAGCCGAGCTGCGAGCTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1117
Db 534 ThrProGlnProProProProProProProProProProProProProPro 541
QY 1118 CCGTGCAGG-----CCAGCGTGGGACATGACAGAGCTGTGAGGCTCTTGTG 1171
Db 542 ProProAlaProProProProProProProProProProProProProPro 561
QY 1172 GAGGAGCCCATGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1231
Db 562 ProProProProGlyProProProProProProProProProProProPro 579
QY 1232 CTTGGGAGGCACTCATGAGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1280
Db 580 ProProProProProProProProProProProProProProProProPro 599

```

RESULT 7

US-10-213-509-5

; Sequence 5, Application US/10213509

; Publication No. US20030054485A1

; GENERAL INFORMATION:

; APPLICANT: Weiss, Joseph

; APPLICANT: Scott, Matthew

; TITLE OF INVENTION: JELLY BELLY GENES AND THEIR USES

; FILE REFERENCE: STAN-232

; CURRENT APPLICATION NUMBER: US/10/213,509

; CURRENT FILING DATE: 2002-08-06

; PRIOR APPLICATION NUMBER: 63/311,720

; PRIOR FILING DATE: 2001-08-09

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 4123

; TYPE: PRT

; ORGANISM: H. sapiens

; US-10-213-509-5

Alignment Scores:

Pred. No.:	0.0114	Length:	4123
Score:	154.50	Matches:	116
Percent Similarity:	28.03%	Conservative:	25
Best Local Similarity:	23.06%	Mismatches:	144
Query Match:	6.71%	Indels:	219
DB:	15	Gaps:	30

US-09-735-251-3 (1-1293) x US-10-213-509-5 (1-4123)

```

QY 196 CCAGAGTCTGTCCAGACCCCTGAGCTGATGATCTACTGTGTGGAAGTAA 255
Db 2773 ProGlyGlyCysProLeuProCysGly-----TrpSer----- 2783
QY 256 AGATGACCAACACTTCACTTCTATGGCATTCAACCTGGGTCCACTGCTCCATG 315
Db 2784 -----AlaTrp-----SerSerTrpAlaProCysAspArgSer--- 2794
QY 316 AAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375
Db 2795 -----CysGly-----SerGlyValArgAla 2801
QY 376 AGAGTTCG-----GGTTTCCACACTGCTCCCTGCGACAG 408
Db 2802 ArgPheArgSerProSerAsnProAlaAlaTrpGlyAlaProCysGlyAsp 2821
QY 409 CAGTCTCTCTTACAGGA-----GGCGTCTTTAAGATGCTCAGCAATAA 453
Db 2822 ArgGlnGluLeuGlnGlyCysHisThrValCysGlyThr----- 2834
QY 454 GGAGCTGTGATCAGATCATCTGCGGACCCAGGCTCAGCAGTGACCTATTG--- 510
Db 2835 GlyIleAlaLysSerLeuGlyAlaGlyValPro-----ProSerSerGlnPheCysThr 2853
QY 511 -----TCTTGGGGT----- 519
Db 2854 LeuArgThrHisGlyMetGlyProThrAspHisSerThrTrpGlyIleGluValPheGly 2873
QY 520 -----TCTCCAGGACCAAGGA 534
Db 2874 TrpThrProTrpThrSerTrpSerSerCysSerGlnSerCysLeuAlaProGlyGly 2893
QY 535 CCT-----CTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 576
Db 2894 ProGlyTrpArgSerArgSerArgLeuCys----- 2903
QY 577 GGTGCTGCTCAGCCAGCCCTGTCATATGCCATATGCTGCTGCTGCTGCTGCTGCT 636
Db 2904 -----ProSerProGlyAspSerSerCysProGlyAspAlaThrGlnGlu 2919
QY 637 CAGTGCCC-----CAATGCTGGGACTCACTCTCTTCCCGGAGCATGCCCTCCA 686
Db 2920 ProCysSerProProIleGluCysThrGlyPheCysAlaProGlyCysThrCysProPro 2939
QY 686 ----- 686
Db 2940 GlyLeuHisLeuHisAsnAlaSerCysLeuProArgSerGlnCysProCysGlnLeuHis 2959
QY 687 -----GCTCAACCGGGATATGCCAGGTGGCTTCTCTGT 719

```

```

; SOFTWARE: Patentin version 3.0
; SEQ ID NO 47
; LENGTH: 624
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-146-473-47

Alignment Scores:
Pred. No.: 0 0112 Length: 624
Score: 152.56 Matches: 97
Percent Similarity: 34.04% Conservative: 70
Best Local Similarity: 19.76% Mismatches: 177
Query Match: 6.62% Indels: 147
DB: 15 Gaps: 19

US-09-735-251-3 (1-1273) x US-10-146-473-47 (1-624)
QY 80 GTGTACACCCACCTTACTCCAAAGTCTATT-----CTTCGTTCCACAGACAGAA 130
DB 24 AlaAlaAlaProAlaGluProLysIleLysValThrValLysThrProLys---Glu 42
QY 131 CTGGGAGATATCTCGCTA---GGGGCTATAGTATTTTCATTTCAGACGCTTATTCT 187
DB 43 LysGluGluPheAlaValProGluAsnSerValGlnGlnPheLysGluAlaIleSer 62
QY 188 GGCACAACTCCAGAGCTGTCTCCAGACCCCTGAGCTGATTGATCTCTACTGTGCTCG 247
DB 63 LysArgPheLysSerGlnThr-----AspGlnLeuValLeuIlePheAlaGlyLys 79
QY 248 AGCTAAAGATGACACAGACACTTGCCTCTATGACATTCACCTGGGTCCACTTCAT 307
DB 80 IleLeuLysAspGlnAspThrLeuIleGlnHisGlyIleHisAspGlyLeuThrValHis 99
QY 308 GTCTCGCAAGCTCTGGCTCAACTGAT---CAGAAACCGGAACCTGTGCACAAAGTG 364
DB 100 LeuValLysSerGlnAsnArgProGlnGlnSerThrGlnProSerAsnAlaAla 119
QY 365 GCTGCCATAGAGAGTTCCGGGTGTGCACACTGCCCTGCACAGCAGCTCTCTTACAGG 424
DB 120 Gly-----ThrAsnThrSerAlaSerThrProArg 130
QY 425 GAGCGCTTTTAAGATGCTCAGCAAT----- 451
DB 131 SerAsnSerThrProIleSerThrAsnSerAsnProPheGlyLeuGlySerLeuGlyGly 150
QY 452 -----AAGCAGTCT 460
DB 151 LeuAlaGlySerSerSerLeuGlyLeuSerSerThrAsnPheSerGluLeuGlnSerGln 170
QY 461 CTGGATAGATATTTGTGCCACCCCA-----GGCCTCAGCAGTACCCCTATT 508
DB 171 MetGlnGlnGlnLeuMetAlaSerProGluMetMetIleGlnIleMetGluAsnProPhe 190
QY 509 GCTCTTGGGGTTCTCCAGACAGGACCTCTCTCT-----GTCTTCGTCATCCCAAT 562
DB 191 ValGlnSerSerLeuSerAsnProAspLeuMetArgGlnLeuIleMetAlaAsnProGln 210
QY 563 ATGCTTGAACGTTGTGTGCTGCTCACCAGCCCTCGTCAATGCC----- 607
DB 211 Met---GlnGlnLeuIleGlnArgAsnProGluIleSerHisLeuLeuAsnAsnProAsp 229
QY 608 ATTGTCTGTTCTGCACTCCCGTAGCAGGACGTCGCCCAATCCCTGGGACTGACTCTCT 667
DB 230 IleMetArgGlnThrLeuGluIleAlaArgAsnProAlaMetMetGlnGluMetMetArg 249
QY 668 TCCCGGAGGATGCCCTCCAGCTCATACCGGGATATGCCAGGTGGGCTTC---CTGTTTCAA 724
DB 250 AsnGlnAspLeuAlaLeuSerAsnLeuGluSerIleProGlyGlyIleThrAsnAlaLeuArg 269
QY 725 GGGCTCTCAGATGATGAGGATGACTTTCACCAACACACAGGTCACACCCCTCTAGCAGT 784
DB 270 ArgMetTyThrAspIleGlnGlnProMetLeuAsnAlaAlaGlnGlnPheGlyGly 289

```

RESULT 8

```

US-10-146-473-47
Sequence 47, Application US/10146473
Publication No. US2003010888A1
GENERAL INFORMATION:
APPLICANT: Scanlan, Matthew
APPLICANT: Gout, Ivan
APPLICANT: Stockert, Elisabeth
APPLICANT: Gure, Ali
APPLICANT: Chen, Yao-Tseng
APPLICANT: Old, Lloyd
TITLE OF INVENTION: Breast Cancer Antigens
FILE REFERENCE: L00461/70130(JRV)
CURRENT APPLICATION NUMBER: US/10146473
CURRENT FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: US 60/291 156
PRIOR FILING DATE: 2001-05-15
NUMBER OF SEQ ID NOS: 82

```


Alignment Scores:	length:	5.64
Fred. No.:	0.023	

```
Db 404 rAlaProLeuProGlnProGlnProAlaSerAlaProGlySerAlaProGlyProValGlnXc 424
Qy 843 -----TCACCCAGAGTGGCTGG----- 860
Db 424 tProLeuProAspAlaThrProSerArgSerPheThrMetSerAspThrGlySerPh 444
Qy 861 -----CCACCGCTTGGCCCTGCCCAGACCTGGGAGACCACTCTCTCACACCCCACT 913
Db 444 eHisLeuProProArgProSerThrProValArg----- 456
Qy 914 CTGGCACCAGGGTCACTCTCTCAAGACCTCAAACTGCTCTGGTTCACAGTCAGGG 973
Db 457 -----ProGlnProLeuProGlnArgPro----- 464
Qy 974 AGCCCATCACCAATGATCTTTAGCCAAAGCC-----TACAGCATGCCCTTCAG 1024
Db 465 -----ProThrProAlaSerValAlaLeuProSerProAlaThrAlaProHisLeuGl 483
Qy 1025 GCTCTGGGAGCCAGCCCTCAAGAGCCAGTGGAGCCAGCCAGCTGCAGCACTACGTGAC 1084
Db 483 uProThrArgValProSerProAlaGProSerAlaAlaProAlaAlaSerValGlnHisH 503
Qy 1085 ATGGCATCCAGGAGCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1144
Db 503 sAlaThrAlaSerAlaGlyAlaHisThrAlaProAlaAlaGlnValHisThrArg 521
Qy 1145 AAGCAGCCCTGGAGCTGATCTTTCTGGAGGAGCCCATGAATCCCTGCTTCCCTGAA 1204
Db 522 ProGln-----AlaThrArgArgArgArgArgArgArgArgArgArgProGly 537
Qy 1205 CCCCAGCA 1213
Db 538 ProProAla 540
```

RESULT 10
US-09-893-519A-14
Sequence 14, Application US/0989351A
Publication No. US20030027243A1
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMESON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BUCMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thara
APPLICANT: HARRIS, Sandra
APPLICANT: KOVARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAQ, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOSITIONS AND METHODS OF USE
FILE REFERENCE: 0342/10548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
PRIORITY FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87

PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(1023)
US-09-893-519A-14

Alignment Scores:
Pred. No.: C.0257 Length: 1023
Score: 148.00 Matches: 91
Percent Similarity: 31.21% Conservative: 17
Best Local Similarity: 26.30% Mismatches: 134
Query Match: 6.43% Indels: 104
DB: 11 Gaps: 15

US-09-735-251-3 (1-1273) x US-09-893-519A-14 (1-1023)

Qy 355 GGACAAAGTGGCTGCCATGAGAGAGTTCGGGTGTTGCACACTGCCCTGCACAGCAGCTC 414
Db 74 GlyProAlaAlaProAlaGluGlyAlaProGlyAlaAlaProGluProProAlaGly 93
Qy 415 CTCTTACAGGGAGGGCTCTTTAAGATGCTCAGCAATAAGGAGTCTCTGGATCAGATCAT 474
Db 94 ArgAlaArgProGlyGly-----GlyGlyProGlnArg 104
Qy 475 TGTGGCCACCCAGCCCTCAGCAG-----TGACCTATTGCC 510
Db 105 ProGlyProProSerProArgArgProLeuValProAlaGlyProAlaProProAlaAla 124
Qy 511 TCTTGGGTCTCCAGGACCAAGACCTCTTCTCTGCTTCTCGCTGATCCCAATATGCTTGA 570
Db 125 LysLeuArgProProProGluGlySer----- 133
Qy 571 TAGTTGGTGGCTGTCTACCCAGCCCTGCTCAATGCCATTGTC----- 613
Db 134 ---AlaGlyAlaCysAlaProValProAlaAlaAlaValAlaAlaGlyProGluPr 152
Qy 614 -----CTGGTTCTGCACCTCGTAGCAGGAG 639
Db 152 oAlaProAlaGlyProAlaLysProAlaGlyProAlaGlyProAlaAlaAlaArgAlaGlyPr 172
Qy 640 TGCCCAATGCTGGGAGTGAAGTCTCTCTTCCCGGAGCATGCCCTCCAGCTATACCGGA 699
Db 172 oGlyProGlyProGlyProGlyProGlyProGlyProGlyProAlaGlyProGlyAla 192
Qy 700 TATGCCAGGTGGCTTCTCTTTTGAAGGGCTCT---CAGATGATGAGGATGACTTTTCCACC 756
Db 192 LaAlaGlnThr-----LeuAsnGlySerAlaAlaLeuLeuAsnSerHisAlaAla 209
Qy 757 AAACAACA-----GGTCCACACCTCTAGCAGTACTACTCCAGCTC 795
Db 209 LaAlaProAlaValSerLeuValAsnGlyProAlaAlaLeuLeuProLeuProLysP 229
Qy 796 CCGCCAGCTCTCCCTGGGTACAGTGGAGCTGCTGGGCC-----G 837
Db 229 roAlaAlaPro-----GlyThrValIleGlnThrProPropheValGlyAlaAlaAlaP 247
Qy 838 GCCATCACCCAGAGTGAAGTGGCCACCCCTTGGCCCTGGCCAGCACTCCCGAGAGCG 897
Db 247 roProAlaProAlaAlaProSerProProAlaAlaProAlaProAla----- 262
Qy 898 CTCTCACACCCAGCTCTCTGGCAGCCAGGGTCACTCTCCAGGAGCACTCCACCAATGTCCTC 957
Db 263 -----AlaProAlaAlaAlaProProProProProAlaProPro- 275
Qy 958 TGGTGTCCAGTCAGGACCCCACTCAACCAATGATCTCTTTCAGCCCAAGCCCTACACATGC 1017
Db 276 ----AlaThrLeuAlaArgPro-----P 282
Qy 1018 CTTTACGGCTCTGGCCAGCCAGCCAGCTTCAGAGCCAGTGGAGCCAGCTCAGCAGCT 1077
Db 282 roGlyHisProAlaGlyProProThrAlaAlaProAlaAlaValProProProAlaAla 301
Qy 1078 ACGTGATGGGCTATCCAGGACGATGAGCTGAGCTGGCGCTGGCGCTCTGCGGCTCTGCGG 1134

Query Match: 6.32% Indels: 148
DB: 15 Gaps: 21

US-09-735-251-3 (1-1273) x US-10-146-473-50 (1-2971)

QY 77 CTGGCTGACAGCCATTCCTCAGAGTCTATTCTT----- 112
| | | | |
Db 1068 LeuAlaProAlaProAtpProSerSerglyLeuProAlaValLeuAsnProArgPro 1087
| | | | |
QY 113 -----CGTTGCCATACACAGACTGGAGAATACCGCAGGGGTAT 157
| | | | |
Db 1398 ThrLeuThrProGlyArgLeuProThrProThrLeuGlyThrAlaAspAlaPrometPro 1107
| | | | |
QY 158 AGTAATTCATTTCAAGACAGCTATTGCTGGAAACTCCAGAGTGCTGCACAGACCCT 217
| | | | |
Db 1108 ThrProThrLeuValArgProLeuLeu-----LysLeuValHissser ProSerPro 1124
| | | | |
QY 218 GAGCTGATTGATCTCATCTACTGHSCTGGGAJCTAAAGATACACAGACTTGACTCT 277
| | | | |
Db 1125 GluVal-----SerAlaSerAla 1130
| | | | |
QY 278 TATGGCATTCAACCT-----GGTCCACTGTCCATGTCTTCGAAAGTCTCGCT 328
| | | | |
Db 1131 ProGlyAlaAlaProLeuThrIleSerSerProLeuHisVal -ProSerSerLeuPro 1149
| | | | |
QY 329 GAACGTGATCAGAACCGGAACCTTTGACAAAGTGGCTGCLATGAGAGAGTTCGGGGG 388
| | | | |
Db 1150 GlyProAlaSerProMetProLeuProAsnSerSerProLeu----- 1164
| | | | |
QY 389 TTGCACACTGCCGTCACAGAGAGTCTCTCTTAAGGAGGCGCTTTTASATGCTCAGG 448
| | | | |
Db 1165 -----AlaSerProValSerSerThrValSerValProLeuHis 1177
| | | | |
QY 449 AATAAGAGTCTCTGGATCAGATATTGTGSCACCCAGCCCTCACCAATGACCTACT 508
| | | | |
Db 1178 SerSerLeuProfile---SerValProThrLeuProAlaProAlaSerAlaProLeu 1196
| | | | |
QY 509 GCTCTTGGGTTCTCCAGACAAGAGCTCTCTCTGCTCTTGTGATCTGATATGCTT 568
| | | | |
Db 1197 ThriLeProfileSerAlaProLeu----- 1204
| | | | |
QY 569 GATACCTGGTGGCTGCTCCACCAAGCTCTCTAATGCCATGTCTGCTGCACTCT 628
| | | | |
Db 1205 ---ThrValSerAlaSerGlyProAlaLeuThrSerValThrProAlaAlaPro 1223
| | | | |
QY 629 STAGAGCCAGTGTCCCAATGCTTGAATTAAGCTTTTGGAGAACTCTCAG 688
| | | | |
Db 1224 ValValProAlaAlaProGlyProProSerHisProSerGlyAlaSerProSerAla 1243
| | | | |
QY 689 TCATACCCGGATACCCAGTGGATTTGCTTTAAGGCGCTTAGATGAGGATGAC 748
| | | | |
Db 1244 SerAlaLeuThrLeu---GlyLeuAlaThrAlaProSerLeuSerSerGlnThrPro 1262
| | | | |
QY 749 TTTACACCA-----AACACACAGCTCC 769
| | | | |
Db 1263 GlyHisProLeuLeuAlaProThrSerSerHisValProGlyLeuAsnSerThrVal 1282
| | | | |
QY 770 ACACCTCTAGCAGT-----ACTCCAGTTCGGCCAGCTCTCCCTGGGTACAGT 820
| | | | |
Db 1283 AlaProAlaCySerProValLeuValProAlaSerAlaLeuAlaSer -ProPhePro 1301
| | | | |
QY 821 GGAGCTGTGGCCCCCGCCCATCACCAAGT----- 853
| | | | |
Db 1302 SerAlaProAsnProAlaProAlaHisAlaSerLeuLeuAlaProAlaSerSerAlaSer 1321
| | | | |
QY 854 ---GAGCTGGCCACCCCTTGCCCTGGCCCAATCTCCGAGAGAGCTCTCCACACCC 910
| | | | |
Db 1322 GlnAlaLeuAlaThrProLeuAlaProMetAlaAlaProGlnThrAlaAlaLeuAlaPro 1341
| | | | |
QY 911 ACTCTGGCACCCAGGGCTATTCTCT-----CAGGAGCTCACCAATGTCC 955
| | | | |
Db 1342 SerPro-AlaProProAlaProLeuProValLeuAlaProSerProGlyAlaAlaPro 1361
| | | | |

185	QY	185	GCTGGCAAACTCCAGGAGTCTGTGTTCCAGACCCCTGAGCTGATTGATCTGATCTACTGCTGT	244
66	Db	66	SerLysArgPheLysSerHis.sThr.....AspGlnLeuValLeuLeuPheAlaGly	82
245	QY	245	CGGAAGCTAAAGATGACACAGACACTTGACTTTATGGCATTCACCTGCTCCACTGTC	304
93	Db	93	LysLeuLysAspGlnAspThrLeuSerGlnHisGlyLeHisAspGlyLeuThrVal	102
305	QY	305	CATCTTCTGGCAAAAGTCTGGCCCAATCTCATCAGAAACGGAACTGTCGACAAAGTG	364
103	Db	103	HisLeuValLeuLysThrGlnAsArgProHisasp	114
365	QY	365	GCTGCCATGACAGAGTCTGGGCTGTTTTCACATTCGCTGTCACACAGGCTCTCTATAC	424
115	Dt	115	HisSerAlaGlnGlnHisThrAsnThrAlaGlySer	125
425	QY	425	GAGCGGTCTTTAAGATGCTCAGTAATAAGGAGTCTCTGGATCAGATCATTTGGCCAC	484
126	Db	126	AsnValThrThrSerSerThrPrAsuSerAsuSerThrSerGlySerAlaThrSerAsn	145
485	QY	485	CCCA.....GGCTCCACAGCT	499
146	Db	146	ProPheGlyLeuGlyGlyLeuGlyLeuAlaGlyLeuSerLeuLeuAsnThr	165
500	QY	500GACCTATT	508
166	Db	166	ThrAsnPheSerGluLeuLysSerHisMetGlnArgGlnLeuLeuSerAsnProGluMet	185
509	QY	509	GCTTTGGGTTCTCCAGGACAAAGACTCTTTCTGTCTCTCTGATGACCAATATGCT	568
186	Db	186	MetValGlnIleMetGluAspProValGlnSerMetLeuSerAsnAspLeuMet	205
569	QY	569	GATACGTTGTCGCTCTTACCTACCAATGCTGTAATGCCAT	610
206	Dt	206	ArgGlnLeuIleMetAlaAspPrHisMetGlnIleLeuIleHisAsnAsnProGluLe	225
611	QY	611GTCCTGGTCTG	622
226	Dt	226	SerHisMetLeuAsnAsnProAsiGluMetAlaGlnThrLeuGluLeuAlaAsnAsnPro	245
623	QY	623CACTCCCTAGCAGGATGTCGCCCATG	649
246	Dt	246	AlaMetMetGlnGluMetValAlaAsnGlnAspArgAlaLeuSerAsnLeuGluSerIle	265
650	QY	650	CCTGGGACTGACTCTCTCTCTGCAAGATGCTGCTG	688
266	Dt	266	ProGlyGlyThrAlaLeuAlaLeuAlaMetGlyThrSerLeuHisGlnLeuLeuSer	285
689	QY	689	TCAATACGGGATATGTAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	730
286	Dt	286	AlaAlaGlnGluGlnProAsiGlyLeuAsnPrHisSerAlaSerAlaSerAlaSerThrLeu	305
731	QY	731TGGATCATAGATGCTCTT.....AGCCCAACAGTC	763
306	Dt	306	GlyGluGlySerGlnProSerAlaThrHisAsnArgAspProLeuProAsnPrAla	325
764	QY	764	AGGTCCACACCTCTAGCAGTAATCAATCTTGGCCAGCTGCTG	811
326	Db	326	ProGlnThrSerGlnSerSerSerAlaSerSerGlyThrAlaSerThrValGlyLysThr	345
812	QY	812GGGTAAGTADAGTACTGCGCTGACCC	841
346	Db	346	ThrGlySerThrAlaSerGlyThrSerGlyGlnSerThrThrAlaProAsnLeuValPro	365
842	QY	842ATCCCG	847
366	Dt	366	GlyValGlyAlaSerMetPheAsnThrProGlyMetGlnSerLeuLeuGlnGlnLeuThr	385
848	QY	848	CAGACT.....	853
386	Dt	386	GluAsnProGlnLeuMetGlnAlaMetLeuSerAlaProGlyMetArgMetMetMetGln	405
853	QY	853	853

```

406 SerLeuSerGlnAsnProAspLeuAlaAlaGlnMetMetLeuAsnAsnProLeuPheAla 425
854 -----GAGCTGGCCACCGCTTGGCCCTGGCC 880
426 GlyAsnProGlnLeuGlnGluMetArgGlnGlnLeuProThrPheLeuGlnGlnMet 445
881 AGCACTCGGAGGAGGAGCTCTCACACCGACTCT ----- 916
446 GlnAsnProAspThrLeuSerAlaMetSerAsnProArgAlaMetGlnAlaLeuLeuGln 465
916 ----- 916
466 IleGlnGlnGlyLeuGlnThrLeuAlaThrGluAlaProGlyLeuIleProGlyPheThr 485
917 -----GGCACCGAGGTCAATCTCTCAGGAGCACTCACCAATCTCTCT 958
486 ProGlyLeuGlyAlaLeuGlySerThrGlyGlySerSerGlyThrAsnGlySerAsnAla 505
959 GGTGTCCAGCTCAGGACGCGCC-----ATCACCAATGATCTCTTCAGGCAAGCC 1006
506 ThrProSerGluAsnThrSerProThrAlaGlyThrThrGluProGlyHisGlnGlnPhe 525
1007 CTACAGCATGCCCTTCAGCCCTCTGG-----CAGCCCGACCTTCAGAGC----- 1051
526 IleGlnGlnMetLeuGluAlaLeuAlaGlyValAsnProGlnLeuGlnAsnProGluVal 545
1052 CAGTGGACAGCCCGACCTCAGCAGCTACGTACATGGGCATCCAGGACGATGACTCAGC 1111
546 ArgPheGlnGlnGlnLeuGlnLeuSerAlaMetGlyPheLeuAsnArgGluAlaAsn 565
1112 CTGCG-GGCCCTCAGCGCCACCGTGGGGACATCCCAAGCAGCCCTGGAGCTCATCTTTGCT 1170
566 LeuGlnAlaLeuIleAlaThrGlyGlyAspIleAsnAlaAlaIleGluArgLeuLeu--- 584

```

RESULTS

```

US-10-293-000-2
; Sequence 2, Application US/10293000
; Publication No. US2003C175278A1
; GENERAL INFORMATION:
; APPLICANT: Montecro, Mervyn J.
; APPLICANT: Nah, Alex L.
; APPLICANT: Perry, George
; APPLICANT: Smith, Mark A.
; TITLE OF INVENTION: UBUTOULIN, A PRESENTIN INTERACTOR AND METHODS OF USING SAME
; FILE REFERENCE: 4:15-175
; CURRENT APPLICATION NUMBER: US/10/293,300
; CURRENT FILING DATE: 2003-04-01
; PRIOR APPLICATION NUMBER: US 60/338,549
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.2

```

Alignment Scores:	
Pred. No.:	0.0488
Score:	13.50
Percent Similarity:	31.94%
Best Local Similarity:	18.75%
Query Match:	6.23%
D3:	12
Length:	595
Matches:	108
Conservative:	76
Mismatches:	159
Indels:	234
Gaps:	18

US-09-735-251-3 (1-12-93) X US-10-293-000-2 (1-5-95)

QY 49 GCGGTGAAGCTGGCTGACGAGCCACTCCAAAGTCTATTCTCGTTGCCAGAGACA :27

```

366 GlyValGlyAlaSerMetPheAsnThrProGlyMetGlnSerLeuLeuGlnGlnIleThr 385
848 CAGAGT----- 853
386 GluAsnProGlnLeuMetGlnAsnMetLeuSerAlaProTyrMetArgSerMetMetGln 405
953 ----- 953
406 SerLeuSerGlnAsnProAspLeuAlaAlaGlnMetMetLeuAsnAsnProLeuPheAla 425
854 -----GAGCTGGCCACCGCTTGGCCCTGGCCCTGGCC 880
426 GlyAsnProGlnLeuGlnGlnMetArgGlnGlnLeuProThrPheLeuGlnGlnMet 445
881 AGCACTGGGAGGACGACCTCTCACACCCGACTCCT----- 916
446 GlnAsnProAspThrLeuSerAlaMetSerAsnProArgAlaMetGlnAlaLeuLeuGln 465
916 ----- 916
466 IleGlnGlnGlyLeuGlnThrLeuAlaThrGluAlaProGlyLeuIleProGlyPheThr 485
917 -----GGCACCCAGGGTCATCTCTCAGGGACCTCACCAGTCTCTCT 938
486 ProGlyLeuGlyAlaLeuGlySerThrGlySerSerGlyThrAsnGlySerAsnAla 505
959 GGTCTCCAGTCAGGGGAGGCC-----ATCACCAATGATCTCTTCAGCCCAAGCC 1006
506 ThrProSerGluAsnThrSerProThrAlaGlyThrGluProGlyHisGlnGlnPhe 525
1007 CTACAGCATGCCCTTCAGGCCTCTGGG-----CAGCCCAAGCTTCAGAGC----- 1051
526 IleGlnGlnMetLeuGlnAlaLeuAlaGlyValAsnProGlnLeuGlnAsnProGluVal 545
1052 CAGTGGCAGCCCCAGCTGCAGCAGCTACGTGACATGGGCATCCAGGACGATGAGCTGAGC 1111
546 ArgPheGlnGlnLeuGluGlnLeuSerAlaMetGlyPheLeuAsnArgGluAlaAsn 565
1112 CTGGC-GCCCTGCAGGCCACCGTGGGACATCCAGCAGCCCTGGAG 1158
566 LeuGlnAlaLeuIleAlaThrGlyGlyAspIleAsnAlaIleGlu 581

RESULT 15
US-10-017-161-1350
; Sequence 1350, Application US/10017161
; Publication No. US20090143669A1
; GENERAL INFORMATION
; APPLICANT: SUZUKI, MAKIKO
; APPLICANT: ASAI, FUYOCHI
; AFFILIANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 08435/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1350
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (21..135)
; OTHER INFORMATION: Variable amino acid
US-10-017-161-1350

Alignment Scores:
Pred. No.: 0.0552 Length: 737
Score: 143.00 Matches: 86
Percent Similarity: 33.73% Conservative: 27

```

```

RESULT 15
US-10-017-161-1350
? Sequence 1350, Application US/10017161
? Publication No. US20101413669A;
? GENERAL INFORMATION
? APPLICANT: SUMA, MAKIKO
? APPLICANT: ASAI, KIYOSHI
? APPLICANT: AKIYAMA, YUTAKA
? APPLICANT: ABURATANI, HIROYUKI
? TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
? FILE REFERENCE: 084335/0152
? CURRENT APPLICATION NUMBER: US/10/017.161
? CURRENT FILING DATE: 2002-12-18
? PRIOR APPLICATION NUMBER: JP 2001/246789
? PRIOR FILING DATE: 2001-06-18
? NUMBER OF SEQ ID NOS: 2433
? SOFTWARE: Patentin Ver. 2.1
? SEQ ID NO 1350
? LENGTH: 737
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (2)...(35)
? OTHER INFORMATION: Variable amino acid
US-10-017-161-1350

Alignment Scores:
Pred. No.: 0.0552
Score: 143.00
Percent Similarity: 33.73%
Conservative: 27

Length: 737
Matches: 86
Conservative: 27

```

Best Local Similarity: 25.67% Mismatches: 86
 Query Match: 6.21% Indels: 136
 DB: 12 Gaps: 17

Search completed: November 12, 2003, 07:00:40
 JCB time : 64.5 secs

US-09-735-251-3 (1-1273) x US-10-017-151-1350 (1-727)

```

QY 326 CQTGAACCTGATCAGAAACGGGAAATTTTGGAAAGTGGTGGCATGAGAGA-----G 379
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 ProGlnProArgGluProProSerPro-TripSerProSerValPro-GlyAsnProSerV 152
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 380 TTCCGGGTGTTCCAGACTCCCTCCACAGCACTCTCTTACAGGAGAGGCTTTTAAAG 439
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 152 alProG.YThrProSerIlePro----- 159
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 440 ATGCTCAGCAATAGAGAGTCTCTGATCAGATATTGGGGACGCC-----A 487
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 160 -----GlyThrProHisSerGlyAsnProSerValProGlyT 172
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 488 GGCCTCAGCAGTGACCTATTCTTTGSGGT----- 519
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 172 hrProGln-----SerTyrLysProLeuSerProGlyAsnProSerValA 187
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 -----TCTCAGGACAAAGACCTCTC----- 541
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 187 rgLysProLeuSerProG.Y.LysProLeuSerProArgLysProLeuSerProGlyAsn 206
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 542 -- TCTGCTTCGGTCATGCAATATGCTTGATAGTTCGTCGCTGCTTACCCAGCCCTC 598
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 207 ProSerValProG.YThrProGlnHisGlyAsnProSerValProGlyAsnPro----- 224
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 599 GTCAATGCCATTGTCTCTCTTCTTCTGCTAGCAGCACTGACCCCAATCTCTGGGACT 658
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 225 -----SerValProGlyAsnProSerValProGlyThr 235
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 659 GACTCTCTTCCGGAGGATGAGTCCGCTGCTATACGGGATATGCCAGCTGGCTCTCTG 718
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 ---ProGlnSerArgGluProGlnSerGlyAsnProSerValProGly----- 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 719 TTTGAAGGGCTCTCAGATGATGATGATGATTTACTTAAATCAATCAGGACCAACCCCT 778
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 252 -----ThrProGlnSerGlnIleThrProGln 260
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 779 ASCAGTACTCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 838
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 261 SerArgLysProLeuSerProArgLysProLeuSerProArgLysProLeuGlyProG 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 839 C-----CATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 883
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 280 yAsnProLeuSerGlnIleThrProGlnHisGlyAsnProSerValProGlyAsnPro 300
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 884 ACTCGGAGAGCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 943
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 300 rValProArgAspProLeuSer-----ProGlyAsnPro 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 944 TCACCAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 997
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 311 o-----SerAlaProG.YThrProGlnSerLeuGluProLeuSerPro 325
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 998 AGCCAAAGCCCTACAGC-----ATGCCCTCAAGCTCTGCGCAGCCAGCTTCAGAGC 1051
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 325 oGlyLysProLeuSerProGlyAsnProPheAsnPro-----GlyAsnProSerValArg 344
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1052 CAGTGGCAGCCCGCAGCTGCGAGCACTGCTGATGGGATCCAGGACGATGAGCTGAGC 1111
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 344 sProLeuSerPro----- 348
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1112 CTGGGCGCCCTGCGAGCCACCGGTGGGACATCAGAGCCCTGGAGCTCATCTTTGCTG 1171
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 349 -----GlyAsnProLeuSerProGlyLysProLeu----- 358
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1172 GAGGAGCCCATGAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1208
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 359 ----SerProArgAsnProSerValProGlyAsnPro 369
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

COM nucleic - protein search, using frame plus 120 model:

Run on: November 12, 2003 06:44:29 ; Search time 18.5 Seconds
(without alignments)
5827.889 Million CpG updates

Title: US-99-735-251-;

Perfect score:

Sequence: 1 GCGAGAGGCCAGAGGAGA/ AGCTTCATATCTCTCTGTC :273

scoring table: BLOSUM62

```
Xqapop 10.9 , Xgapex= 0.5
```

ygapop 10.0 , ygapex= 0.5
 ygapop 10.0 , ygapex= 0.5

Gapop 6.0
Gapex+

Delop 6.0, Delox 7.0

Searched: 328717 seqs, 423:0853 residues

Total number of hits satisfying chosen parameter-rs: 657434

Minimum 23 seq length: 0

Min: num	DB	seq	length: 0
Max: num	DB	seq	length: 2000000000

Post-processing: Minimum Max-Cut &

Post-processing: Minimum March 1000
Maximum March 1000

Maximum Match: 100%
Listing first: 45 slides

Command line parameters:

```
-MODEL=frame+ n2n model -DEV=x:n
```

```
-MODES=frame+ h2p.model -DEV=X,p
-Q=/cqn2 1/USPTO s000] p/US9735251/rupa: 12112003 061636 28180/app mervy fasta 14:5
```

-CB=Issued Patents AA -OFMT=faster -SUFFIX=X -VINMATCH=C -LOOPL=0

```

-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blocks:m62 TRANS=human:40.03:

```

```

DOO:MAX=0  UNITS=DIGS  CIGARS=1  CIGARS=MAX  CIGARS=MAX
- LIST=45  -DOCALIGN=200  -THR SCORE=900  -THR VAX 100  -THR MIN=0  -ALIGN=15

```

```
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
```

```

-USER=US09735251 @CGN 1 1 27 @runat : 212003 061636 28:80
-NCPU=6 -ICPU=3

```

```
-NC MMAP -LARGEQUERY -NEG_SCORES=0 -WAT -DSPBLOCK=100 -LONGLOG
```

```
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
```

```
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DEF.CP=6 -DEF.EXT=7
```

Database : Issued Parents AA:*

```
1: /cqn2_6/ptcdata/1/ind/SA_CMB.pcp:
```

2: /cgn2_6/prodata/1-14114-0-03-000-

3: /cgn2_6/p/ptdata/ + 1000000000

4: /cgn2_e/product=1 -r 2 -v 3 -c 3 -c 3

```
5: /cgn2_6/p/ptdata/ -d /cgn2_6/p/ptdata/
```

6: /cgn2_6/podestati_13a/back:iesi.pdf

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2133	92.7	410	3	US-08-728-758A-4	Sequence 4, Appl 1
2	2129	92.5	426	3	US-08-728-758A-4	Sequence 2, Appl 1
3	163	71.1	816	4	US-09-266-225C-12	Sequence 12, Appl 1
4	145.5	6.3	2972	4	US-09-579-181-2	Sequence 2, Appl 1
5	145.5	6.3	3118	4	US-09-579-181-2	Sequence 1, Appl 1
6	143	6.1	2294	4	US-09-253-991A-17231	Sequence 17231, A
7	142.5	6.2	1093	4	US-09-253-991A-17231	Sequence 17231, A
8	142	6.2	902	1	US-08-396-475B-6	Sequence 21827, A
9	142	6.2	902	1	US-08-818-823-6	Sequence 6, Appl 1
10	140.5	6.1	1060	4	US-08-911-393-2	Sequence 6, Appl 1
11	139.5	6.1	1093	3	US-08-548-860D-55	Sequence 2, Appl 1
12	139.5	6.1	1093	5	PCF-0394-04496-55	Sequence 55, Appl 1

SEQUENCE CHARACTERISTICS:
 LENGTH: 410 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 US-08-725-758A-4

Alignment Scores:

Pred. No.: 1,42e 182 Length: 410
 Score: 2133.00 Matches: 410
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 92.66% Indels: 0
 D1: 3 Gaps: 0

US-09-735-251-3 (1-1273) x US-08-725-758A-4 (1-410):

```

QY 44 ATGCTCTCTCAGACGGACCTTGGCGTGAAGCTGGCTGACGAGCCATTACTCTCAAG 103
DB 1 MetSerLeuSerAspTrpHisLeuAlaValLysLeuAlaAspGlnProLeuTrpOlys 20
QY 104 TCTATTCTTCGTTGCCAGACAGCACTGGGAGAAATACCTGGTAGGGGCTATATATT 163
DB 21 SerLeuArgLeuProGluTrpGluLeuGlyGlnTyrSerLeuLeuGlyGlyTyrSerLeu 40
QY 164 TCATTCTCAAGCAGCTTATTGCGGCAACACCCAGGAGCTGTTTCCCAACCTGAGCG 223
DB 41 SerPheLeuLysGlnLeuLeuAlaGlyLeuLeuGlnSerValProAspProGluLeu 60
QY 224 ATTGATCTGATCTACTGTGTGGAAGCTAAAGATGACGAGACCTTACTTCTATGCG 283
DB 61 IleAspLeuIleTyrCysGlyAlaLysLeuLysAspAspGlnTrpLeuAspPheTyrGly 80
QY 284 ATTCACCTGGTCCACTGCTCCGCTGCTGGGCAACTCTGGCTGCTGCTGCTGCTGCTG 343
DB 81 IleGlnProGlySerThrValHisValLeuArgLysSerTrpProGluProAspGlnLys 100
QY 344 CCGGACCTGTGCACAAAGTGTGTCATGACGAGAGCTCCGGTGTGTGACACTGCGCTG 403
DB 101 ProGluProValAspLysValAlaAlaValArgGluPheArgValLeuHisThrAlaLeu 120
QY 404 CACGACGCTCTCTCTACAGGAAGCGGCTTTAAGATGCTCAGCAATAGAGAGTCTGTG 463
DB 121 HisSerSerSerTyrArgGlyAlaValProLysMetLeuSerAsnLysGlnSerLeu 140
QY 464 GATCAGATCTCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 523
DB 141 AsnGlnIleLeuValAlaTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
QY 524 CAGGACAGGACCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 583
DB 161 GlnAspLysAspLeuPheSerValPheAlaAspProAsnMetLeuAspSerLeuValPro 180
QY 584 GTCACCCAGCCCTGCTCAATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 643
DB 181 AlaHisProAlaLeuValAsnAlaLeuValLeuValLeuHisSerValAlaGlySerAla 200
QY 644 CCAATGCTGGGAGCTACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
DB 201 ProMetProGlyThrAspSerSerSerArgSerMetProSerSerSerTyrArgAspSer 220
QY 704 CCAAGTGGCTTCTGTTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 763
DB 221 ProGlyGlyPheLeuPheGluGlyLeuSerAspAspGlnAspAspPheHisProAsnThr 240
QY 764 AGGTCCACACCTCTACAGTACCTCCAGCTCCCGCCCAACCTGCTGCTGCTGCTGCTGCTG 823
DB 241 ArgSerThrProSerSerThrProSerArgProAlaSerLeuGlyTyrSerGly 260
QY 824 GCTGCTGGGCCCCGCCCATCACCCAGAGTACGCTGGCCACCGCTTGGCCCTGGCCAGC 883
DB 261 AlaAlaGlyProArgProIleThrGlnSerGlnLeuAlaThrAlaLeuAlaLeuAlaSer 280

```

```

QY 884 ACTCCGAGAGCAGCTCTCAGACACCCAGCTCTCTGGCAGCCAGGCTCATCTCTCAGGACC 943
DB 281 ThrProGlnSerSerHisThrProThrProGlyThrGlnGlyHisSerSerGlyThr 300
QY 944 TCACCAATGCTCTCTGCTGCTCAGTTCAGGACGCCCATCACCATGATCTCTTTCAGCCAA 1003
DB 301 SerProMetSerSerGlyValGlnSerGlyThrProIleThrAsnAspLeuPheSerGln 320
QY 1004 GCCCTACAGCATGCCCTTCAGGCTCTGGGAGCCAGCCCTTCAGAGCCAGTGGCAGCCC 1063
DB 321 AlaLeuGlnHisAlaLeuGlnAlaSerGlyGlnProSerLeuGlnSerGlnTrpGlnPro 340
QY 1064 CAGCTGCAGCAGCTACGTCAGCATCGGCATCCAGGACCATCAGCTGAGCCTCGGCGCCCTGC 1123
DB 341 GlnLeuGlnGlnLeuA-GaspMetGlyLeuGlnAspAspGlnLeuSerLeuArgProCys 360
QY 1124 AGGCATCGTGGGAGCATCCAGACGCCCTCGGAGCTCATCTTTGCTGGAGGAGCCCAT 1183
DB 361 ArgProProValGlyThrSerLysGlnProTrpSerSerSerLeuLeuGluGluProHis 380
QY 1184 GAATCTCTGCTCTCCCTGAAACCCAGCCAGTTCGACAGGCTACTGCGCTTGGGAGGCA 1243
DB 381 GlnLeuProAlaSerProGluProProAlaSerCysArgGlyTyrCysProTrpGluAla 400
QY 1244 CTCATGAGGTGCTCTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1273
DB 401 LeuMetLysValProProSerLeuProVal 410

```

RESULT 2

US-08-725-758A-2
 Sequence 2, Application US/08725758A
 Patent No. 6160108
 GENERAL INFORMATION:
 APPLICANT: Reed, Guy
 APPLICANT: Clerent, Christophe Y.
 TITLE OF INVENTION: NOVEL PLATELET ACTIVATION PROTEIN
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2824
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/725,758A
 FILING DATE: 04-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/005,074
 FILING DATE: 06-OCT-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Fraser, Janis K.
 REGISTRATION NUMBER: 34,819
 REFERENCE/DOCKET NUMBER: 05433/020001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-542-5070
 TELEFAX: 617-542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 426 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 US-08-725-758A-2

Alignment Scores:

Pred. No.: 3,29e-182 Length: 426
 Score: 2129.00 Matches: 429
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 92.48% Indels: 0
 DB: 3 Gaps: 0

US-09-735-251-3 (1-1273) x US-09-735-251-3 (1-426)

QY 44 ATGTCCTCTCAGACTGCGACCTGCGGTGAAGTGGCTGACAGGCGACTTACTCCCAAG 103
 DB 1 MetSerLeuSerAspTrpHisLeuAlaValLysLeuAlaAspGlnProLeuThrProLys 20
 QY 104 TCTATTCCTGGTGGCAGACAGACTGGGAGTAATACTGCTAGCGGCTATAGTATT 163
 DB 21 SerileLeuArgLeuProGlnThrIleuGlyValTyrSerLeuGlyGlyTyrSerIle 40
 QY 164 TCATTTCTGAAGCAGCTTATTGCTGCGAACTCCAGAGTCTGTCCAGAGCCCTGAGCTG 223
 DB 41 SerPheLeuLysGlnLeuLeuLeuValLysLeuGlnGluSerValProAspProGlnLeu 50
 QY 224 ATTGATCTGATCTACTGTGGTGGAGGTTAAAGATGACACAGACACTTGTCTTATGCT 283
 DB 61 IleAspLeuLeuTyrCysGlyArgLysLeuLysAspGlnThrLeuAspPheTyrGly 80
 QY 284 ATTCAACCTGGTCTACTGTCCATTTCTGGGAAAGTCTGCTGAACTGATGACGAA 343
 DB 81 IleGlnProGlySerThrValHisValLeuArgLysSerTrpProGlnProLys 100
 QY 144 CCGAACTGTGGGACAAAGTGGTGGCATGAGAGTTCGGGTGTGGCACTGCGCTG 403
 DB 101 ProGlnProValAspLysValAlaValSerArgGluPheArgValLeuHisValLeu 120
 QY 404 CACAGCAGCTCTCTTACAGGAGGCGGTCTTTAAGATGCTCAGCAATAAGAGTCTCTG 463
 DB 121 HisSerSerSerTyrArgGluAlaValPheLysMetLeuSerAsnLysGluSerLeu 140
 QY 464 GATCAGATCATTTGGCCACGCCCAAGCTTCAGTATGACCTATGCTTGGGGTCTG 523
 DB 141 AspGlnIleValAlaThrProLysLeuSerSerAspProIleAlaLeuGlyValLeu 160
 QY 524 CAGCAGAGGACCTCTCTGCTGCTGCTGATCCCATATGCTGTGATAGTGTGCTGCT 583
 DB 161 GlnAspLysAspLeuPheSerValPheAlaAspProAsnMetLeuAspThrLeuValPro 180
 QY 584 GCTCACCAGCCCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 643
 DB 191 AlaHisProAlaLeuValAsnAlaValValLeuValLeuHisValAlaGlySerVal 200
 QY 644 CCAATGCTGGGATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
 DB 201 ProMetProGlyThrAspSerSerArgSerMetProSerSerTyrArgAspMet 220
 QY 704 CCAGTGGCTCTCTGTTGAAGGCTCTCAGCATGATGAGGATGATTCACCAACACG 763
 DB 221 ProGlyGlyPheLeuPheGluGlyLeuSerAspGlnAspPheHisProAsnThr 240
 QY 764 AGGTCCACACCTCTAGCAGTACTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 823
 DB 241 ArgSerThrProSerSerSerThrProSerSerArgProAlaLeuLeuGlyTyrSerGly 260
 QY 824 GCTGCTGGGCGCGCGCCCATCACCAGAGTATGTCGACAGGCTTGGGCTGGCAG 883
 DB 261 AlaAlaGlyProArgProIleThrLysSerGluLeuAlaThrAlaLeuAlaLeuAlaSer 280
 QY 884 ACTCGGAGAGCAGCTCTCAGACAAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 943
 DB 281 ThrProGluSerSerSerHisThrProThrProGlyThrGlnGlyHisSerSerGlyThr 300
 QY 944 TCACCAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1003
 DB 301 SerProMetSerSerGlyValGlnMetGlyThrIleThrAsnAspLeuPheSerGln 320

QY 1004 GCCTTACAGATGCGCTCTGAGGCTCTGAGGAGCCAGGCTTTCAGAGCAGTGGGAGCCC 1063
 DB 321 AlaLeuGlnHisAlaLeuGlnAlaSerGlyGlnProSerLeuGlnSerGlnTrpGlnPro 340
 QY 1064 CAGCTCGACAGCTACGTCACATGCGCATCCAGGACGATGAGCTGAGCTGGGCGCTGC 1123
 DB 341 GlnLeuGlnGlnLeuArgAspMetGlyIleGlnAspPheGluLeuSerLeuArgProCys 360
 QY 1124 AGCCACCCGCTGGGACATCCAGCAGCCCTGGAGCTCATCTTTGCTGAGGAGCCCAT 1183
 DB 361 ArgProProValGlyThrSerLysGlnProTrpSerSerSerLeuLeuGluGluProHis 380
 QY 1184 GAATCTCTCTGCT 1243
 DB 381 GluLeuProAlaSerProGlnProProAlaSerCysArgGlyTyrCysProTrpGluAla 400
 QY 1244 CTCATCAAGGTGCT 1270
 DB 401 LeuMetLysValProProSerLeuPro 409

RESULT 3
 US-09-266-225D-12
 ; Sequence 12, Application US/09266225D
 ; Patent No. 6573364
 ; GENERAL INFORMATION:
 ; APPLICANT: Nandabalan, Krishan
 ; APPLICANT: Kingsmore, Stephen
 ; APPLICANT: Tchernev, Velizar
 ; TITLE OF INVENTION: Isolation and Characterization of Hermansky-Pudlak
 ; TITLE OF INVENTION: Syndrome (HPS) Protein Complexes and HPS Protein-
 ; TITLE OF INVENTION: Interacting Proteins
 ; FILE REFERENCE: 15966-523
 ; CURRENT APPLICATION NUMBER: US/09/266,225D
 ; CURRENT FILING DATE: 1999-03-10
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 12
 ; LENGTH: 816
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-266-225D-12

Alignment Scores:
 Pred. No.: 3,64e-06 Length: 816
 Score: 163.00 Matches: 88
 Percent Similarity: 36.48% Conservative: 28
 Best Local Similarity: 27.67% Mismatches: 125
 Query Match: 7.08% Indels: 77
 DB: 4 Gaps: 11

US-09-735-251-3 (1-1273) x US-09-266-225D-12 (1-816)

QY 311 CTGCGAAAGTCTCTGGGCTCGAACCTGATCAGAAACCGGAACCTGTGCAAAAGTGGCTGCC 370
 DB 491 LeuArgAspGlyProSerAlaProLeuGluAlaProGluProArgLysProValThrAla 510
 QY 371 -----ATGAGAGATTCCGGGTGTTGCACACTGCGCTGCACAGCAGCTCTCTTAC 421
 DB 511 GlnGluArgSerHisArgGluGluLysArgArgArgGlnGluArgAlaLysGlu 530
 QY 422 AGGAGGCGCTCTTTAAGATGCTCAGCAATAGGAGTCTCTGATCATGATCATTTGTGGCC 481
 DB 531 ArgGlu-----LysArgArgGlnGluArgGluArgLysGluArgGlyAlaGlyAla 547
 QY 482 ACCCAGGCGCTCAGCAGTGCACCTATTGCTTGGGTCTTCAGGACAAGACCTCTTC 541
 DB 548 SerGlyGlyProSerThrAspProLeuAlaGlyLeuValLeuSerAspAsnAsp-ArgSe 567
 QY 542 TCTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
 DB 567 rLeu-----LeuGluArgTrpThrArgMetAlaArgProAlaAl 580

D6	612GluProThrGlyProGlnProGlnIle	620
QY	782	AGTA-----CCTCCAGCTCGGCCAGCTCTCCGTGGTACACTGA	823
D6	620	rAlaGlySerThrSerGlyProValProLlnProAlaOysProPro	635
QY	824	GCTGCTGGCGCCGGCCATCACAGAGTGAGCTGGCCACG---CGTTGGCCCTGGTC	880
D6	636	-----ProGlyProAlaPheHisProThrGlyProGlyProLeuProValHr	652
QY	881	AGCACTCCGAGAGCAGCTCTCATACAGCACTCTCTGGCACCCAGCGGTATCTCTTAGCG	940
D6	652	oAlaProProGlnIleAlaThrSerThrSerLeuLeuAlaAaGln-----	667
QY	941	ACCTACCATTGCTCTCTGCTGTCTAGTCAGGAGCCGCATCACCAAATCATCTCTTCACG	1000
D6	668	-----SerLeuValProProProGlyLeuPro-----GlySerS-679	679
QY	1001	CAAGCCCTACAGCATGCCCCCTCTCTGCGACCCAGAGCCCTTCAGAGCCAGTAGCAG	1060
D6	679	rThrProGlyValLeuProTyrlleProProGlyLeuProProProAspAlaGlyAlaI	699
QY	1061	CCCAGCTGCAGCAGCTACCTGATATGCGATCANGACCATCACTCACTCCGCGACG	1120
D6	699	aProGlnSerSerMetSerGln-----	706
QY	1121	TGCAGCCACCGCTGGAGCATCAAGCAAGCTCTAGAGCTATCTTTGTGGAGGAGC	1180
D6	707	SerProAspValAsnLeuValThrGlnGlnLeuSerLysSerGlnAlaGalAspPhe	725
QY	1181	CATTAACCTCCTCTCTCCCTGAAATCCCAAGTAAAGTGGCAGAGCTAC	1228
D6	726	LeuProProValPheSerGlyThrProLysGlySerGlyAlaGlyTy-	747

```

RESULT 4
US-03-579 181-2
; Sequence 2, Application US/09/079,81
; Patent No. 6365472
; GENERAL INFORMATION:
; APPLICANT: Chivria, John
; APPLICANT: Yachuk, Peter
; TITLE OF INVENTION: SNF2 Related Inhibitors : Protease Inhibitors
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/079,81
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/116,420
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.3
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-03-579-181-2

```

```

Alignment Scores:
Pred. No.:      0.000237
Score:          145.50
Percent Similarity: 33.80%
Best Local Similarity: 23.74%
Query Match:    6.32%
DB:             4
Length:         124
Matches:        12
Conservation:    60
Vismatches:     181
Index:          145
Gaps:           2

```

1271	Db		GluVal	-----SerAlaSerAla	1276
278	QY		TATGGCAITCAACCT	-----GGGTCCACTGTCATGTTCTGCGAAAGTCCTGGCCCT	328
1277	Db		ProGlyAlaAlaProLeuThrIleSerSerProLeuHisVal	---ProSerSerLeuPro	1295
329	QY		GAACCTGATCAGAAACCGAACTGTGGACAAAGTGGCTGCCATGAGAGAGTTCCGGGTG	388	
1296	Db		GlyProAlaSerSerProMetProIleProAsnSerSerProLeu	-----	1310
389	QY		TTGCACACTGGCCCTGCACAGCAGCTCTCTTACAGGAGCGCGTCTTTAAGATGCTCAGC	448	
1311	Db		-----AlaSerProValSerSerThrValSerValProLeuSer	1323	
449	QY		ATAAGGAGTCTCTGATCAGATCATTTGTGGCCACCCAGGCGCTCAGCATGACCTATT	508	
1324	Db		SerSerLeuProIleSerValProThrThrLeuProAlaProAlaSerAlaProLeu	1342	
509	QY		GCTCTTGGGTTCTCCAGCACAAGGACCTCTTCTGCTCTTGCTGATCCCAATATGCTT	568	
1343	Db		ThrIleProIleSerAlaProLeu	-----	1350
569	QY		GATACCTGGTGCTGCTCACCCAGCGCTGTCATATGCCATTGCTCGTCTGCACTCC	628	
1351	Db		-----ThrValSerAlaSerGlyProAlaLeuLeuThrSerValThrProLeuAlaPro	1369	
629	QY		GTAGCAGGAGTGCCTCCCAATGCTGGAGTACTCTCTTCCGGAGCATGCCCTCCAGC	688	
1370	Db		ValValProAlaAlaProGlyProProSerLeuGlnProSerGlyAlaSerProSerAla	1389	
689	QY		TCATACCGGATATGCCAGTGCGCTTCTCTTTGAAGGCTCTCAGATGATGAGATGAC	748	
1390	Db		SerAlaLeuThrLeuGlyLeuAlaThrAlaProSerLeuSerSerGlnThrPro	1408	
749	QY		TTTCACGGA-----AACACCCAGGTCC	769	
1409	Db		GlyHisProLeuLeuLeuAlaProThrSerSerHisValProGlyLeuAsnSerThrVal	1428	
770	QY		ACACCTCTAGAGT-----ACTCCAGCTCCGCCACGCTCCCTGGGTACAGT	820	
1429	Db		AlaProAlaCysSerProValLeuValProAlaSerAlaLeuAlaSerProPhePro	1447	
821	QY		GGAGCTGTGGGCCCGGCCCATCCACGAGT-----	853	
1448	Db		SerAlaProAsnProAlaProAlaGlnAlaSerLeuLeuAlaProAlaSerSerAlaSer	1467	
854	QY		---GAGTGGCCACCGCTGGCCCTGGCCAGTCCCGAGAGCAGCTCTCACACACCG	910	
1468	Db		GlnAlaLeuAlaThrProLeuAlaProMetAlaAlaProGlnThrAlaLeuAlaPro	1487	
911	QY		ACTCTGGACCCAGCGGTCTATCTCT-----CAGGACCTCACCATGTCTCC	955	
1488	Db		SerProAlaProLeuAlaProLeuProValLeuAlaProSerProGlyAlaAlaPr	1507	
956	QY		TCTGGTGTCCAGTCAGGAGCGCCCATCACCAATG-----	989	
1507	Db		oAlaLeuAlaSerSerGlnThrProValProValMetAlaProSerSerThrProGlyTh	1527	
990	QY		---ATCTCTTACGCCAAGCCCTACAGCATGCCCTTC-----	1022	
1527	Db		rSerLeuAlaSerAlaSerProValProAlaProThrProValLeuAlaProSerSerTh	1547	
1023	QY		-----AGGCCTCTGGGAGCCCGCCAGCTTTCAGACCCA-	1053	
1547	Db		rGlnThrMetLeuProAlaProValProSerProLeuProSerProAlaSerThrGlnT	1567	
1054	QY		-----GTGGCAGCCCGAGCTGCAGCAGCTACGTGACATGG---CATCCAGGACGATGAG	1105	
1567	Db		hrLeuAlaLeuAlaProAlaLeuAlaProThrLeuGlyGlySerProSerGlnThr	1586	
1106	QY		CTGAGCCTCGGGCCCTGCGAGCCA-----CCGTTGGGAGCATCCAAAGCAGCCC	1153	

Score: 142.50 Matches: 105
 Percent Similarity: 34.23% Conservative: 47
 Best Local Similarity: 23.65% Mismatches: 174
 Query Match: 6.19% Indels: 118
 DB: 4 Gaps: 16

US 09-735-251-3 (1273) X US-09-252 921A-21829 (1-1093)

QY 92 CTTATCGAAGCTCTATTCTGCTGGTTCAGACACAGAACTGAGAGATACTGCTAGGG 151
 DB 441 ValThrProValSerLeuThrArgThrProValProArgThrPheArgAsp 457
 QY 152 GGCATAGTATTTCATTCGAGAGAGATTATCTGCTGCACAGCTCCAGAGCTCTGTTCCA 211
 DB 458 -----PheArgLeuAlaSerSerArgProLeuAlaLeuThrArgProCysTrp 474
 QY 212 GACCTGAGCTGATTCATCTGATCTATCTGCTGGAAGCTTAAAGATGACACACACTT 271
 DB 474 o-----LeuSerSerArgAlaAlaThrArgValMe 484
 QY 272 GACTTCTATGGCATTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 323
 DB 484 tProAlaLeu-----LeuProMetProProLeuArgLeuLeuSerSerThrProValArg 503
 QY 324 -----GGCTGAACCTGATCAGAACCGGAACTGTGGACAAAGTGGCTGCCATG 373
 DB 503 eSerThrAlaProLeuAlaLeuThrProArgLeuLeuSerArgLeuAlaPro 522
 QY 374 AGAGAGTTCGGGTTTCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421
 DB 523 -----CysSerValThrProAlaSerLeuSerSerLeuProProTrpLeuSe 538
 QY 421 -----CCTCTTAC 421
 DB 538 rSerAlaAlaTrpLeuValSerAlaSerAlaProAlaLeuGluGlyValProProArgLe 558
 QY 422 AGGAGGGCGGTCTTAAGATCTCTAATAAAGAGTCTC-----TCGAT 466
 DB 558 uSerArgAlaAlaAlaAlaAlaValArgLeuProSerLeuThrSerAlaProProTrpLe 578
 QY 467 CAGATCATTTGGCCACCGCCAGGCTTACAGATGACCTATCTCTGCTGCTGCTGCTGCT 526
 DB 578 uSerSerThr-----ProProArg-----LeuThrLeuAlaLeuProTrpLeuSerSe 594
 QY 527 GACAGAGACCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 586
 DB 594 rArgProProSerProLeuSerSerSerProProSerArgLeu-----ArgSe 610
 QY 587 CACGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 637
 DB 610 oSerArgProAlaSerThrProLeuAlaThrProLeuAlaThrProLeuAlaThrProLeu 630
 QY 638 AGTSCCCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 697
 DB 630 gProLeuSerProMetThrLeuProProArgLeuSerSerCysSerArgAlaPheThrAl 650
 QY 638 GATATGCCAGCTG-----GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 733
 DB 650 aThrPheGluValLeuGluLeuSerProAlaAlaLeuSerThrCysArgAlaSerIleAl 670
 QY 734 GATGATGAGGATG----- 746
 DB 670 aMetProProPheAlaAlaIleSerGlnAspTrpLeuLeuSerIleValSerAlaAlaThr 590
 QY 747 -----ACTTTCCACCAACACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 799
 DB 690 rPheArgValPheSerLeuThrSerSerProProCysTrpAlaArgLeuProSerValAl 710
 QY 803 C-----CACCTCCCTGGGTACACTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 835
 DB 710 aCysArgSerProLeuAlaAlaIleArgProProProAlaLeuProThrAlaPheAlaAla 730
 QY 836 CGGCCCATCACCAGAGTGAAGTGGCCACCGCTTGGCCCTGGCCAGCACTCCGAGAGC 895

DB 730 gSerArgArgProSerLeuAsnSerLeuProProTrp-----SerLeuArgArgLe 747
 QY 896 AGCTCTTCACACACCA-----CTCCTGGCACCACCGGTCACTTCTCTCA 937
 DB 747 uAlaThrSerThrArgThrProAlaLeuProLeuAlaLeuProProSerArgLeuSerLy 767
 QY 938 GGGAGCTTCACCAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 997
 DB 767 sAlaProProValAlaLeuLysProAlaSerAlaThrSerThrProProArgLeuSerAr 787
 QY 995 AGCCAAGTCTACAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1054
 DB 787 gluAlaArgTrpThrPheAsnSerLeuArgLeuAspSerArgProProAlaArgLeuSe 807
 QY 1055 TGGCAGCTCCACCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1114
 DB 807 rSerCysProThrValThrArgCysAlaProTrpLeuSerArgLeuProSer--SerArg 826
 QY 1115 CGGCTCTGAGCCACCGCTGGGACATCCAGAGCCCTGAGCTCATCTTCTGCTGAG 1174
 DB 827 LeuSerSerArgProProArgLeuThrSerArgProPro-----ArgLeuLeuSer 843
 QY 1175 GAGCCC 1180
 DB 844 ValPro 845

RESULT 8

US-08-396-479B-6
 ; Sequence 6, Application US/08396479B
 ; Patent No. 5612455

GENERAL INFORMATION:

APPLICANT: HOBY, Timothy
 TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HORNBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Ebbacadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/396.479B
 FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-59450-1/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 494-8700
 TELEFAX: (415) 494-8771
 TELEX: 210 277299

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
 LENGTH: 902 amino acids
 TYPE: amino acid
 TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-396-479B-6

Alignment Scores:

Pred. No.: 0.000288 Length: 902
 Score: 142.00 Matches: 91
 Percent Similarity: 35.71% Conservative: 29
 Best Local Similarity: 27.08% Mismatches: 102
 Query Match: 6.17% Indels: 115


```
QY 430 GGTCTTTAAGATGC-----TCAGCAATRA 453
Db 261 rGPrCaAlaAaCyProThrThrProPrCaAlaAsnAlaProAlaSerAlaAlaAla 281
QY 454 GGAGTCTCTGATCAGATCA-----TTGTGSCCACCAGGCTGAGAGTACCC 504
Db 281 snTPrProThrPrAlaProSerProAlaMetProProSerThrSerAlaThrArgL 301
QY 505 TATTGCTCTTGGGTTCC-----TCAGCAACAAAGACCTCTTCTC 543
Db 301 euAAlaAaCySlyHisAspGlyAspAlaGlyArgAlaSerGluThrPThrArgAspT 321
QY 544 TGTCTTCGCTGATCCCAATATGCTTG-----ATACGTTGTTGCTGCTGCTCA 588
Db 321 hrSerProCyValProAsnCysValAsnTrpAlaGlyAlaGlyAlaProThrSerAlaValA 341
QY 589 CCCAGCCTCTGATACCCATGCTTGGTCTGACCTCTGATCCAGGAGTACCCCAAT 648
Db 341 laLysThrSerProTrpPro---ArgTrpProGlyMetProAlaGlyProArgProSerC 360
QY 649 GCCTGGGACTGACTCTCTTCCCGAGCATGCTCCAGCTCAACCCGGATATGCCAGG 728
Db 360 yStrAlaVal-----AlaAlaTrpAlaProAlaAaGlyGlyTyAlaAlaA 376
QY 709 TGGCT-----TCCGTTTGAAGGCTCTCAGATATGAGCATGACTTTTACC----- 755
Db 376 laProArgAsnSerCysSerAlaGlyArgLeuArgArgGValArgAlaSerAlaArgA 396
QY 755 CC-----TCCGAGCCTCTGCGGTACA----- 818
Dl 436 laArgArgThrSerSerAlaProAlaGlyArgAlaThrProAlaSerAspAlaArgAspS 456
QY 819 -----GTGAGCTGTGGGCGCGGCTGATCA-----CCAGAGCTGA 855
Dl 456 rGySHisProValArgLysAspGlyAlaGlyProAlaSerThrGlyArgSerArgArgA 476
QY 856 GCTG-----AlaProArg-----ProAlaAlaSerAlaHisArgProSerA 530
Dl 476 rGTrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 496
QY 861 -----CAGTCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 920
Dl 496 laCysTyPrProProProAlaThrAlaArgAlaGlyThrProAlaAlaArgAlaSerAlaA 516
QY 901 TCACACACGACTCTGCGACCCAGGCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 960
Dl 516 rGThr-----AlaProArg-----ProAlaAlaSerAlaHisArgProSerA 530
QY 961 TGTCCAGTCAGGAGCCCATCACTGATCTCT-----TGTCTCTGAGTCTGAGTCTGAGTCT 995
Dl 530 laSerAlaThrGlyTrpProGlyProAlaLeuAlaCysProAlaAlaGlyArgArgP 550
QY 996 -----TCAGTCAGTCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 1035
Dl 550 toSerProAlaProAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 570
QY 1036 GCCCAGCCTTCAGAGCCAGTGGCAAGCCAGCTCAGCAGTATGTGATGAGTGGCATCCA 1095
Dl 570 rGProGlyArgSerProAlaThrAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 587
QY 1096 GGACGATGAGTCTGAGCCTGCGGCTTGCAGGCTACCGTGGGACATCCAGAGAGCCCTG 1155
Dl 588 -----ProThrProAlaArgProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 600
QY 1156 GAGCTCATCTTTGCTGGAGGAGCCATGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1215
```

```
Db 601 -----ProSerAlaProArgProAlaArg 609
QY 1216 TTGCAGAGGCTACTGCCCTTGGAGGCA 1243
Dl 609 gArgArgValArgTrpProTrpArgSer 618
RESULT 14
US-09-252-991A-30052
; Sequence 30052, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,798
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30052
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30052
Alignment Scores:
Pred. No.: 0.00896 Length: 719
Score: 136.00 Matches: 118
Percent Similarity: 30.92% Conservative: 36
Best Local Similarity: 23.69% Mismatches: 180
Query Match: 5.72% Indels: 164
Gaps: 26
US-09-735-251-3 (1-1273) x US-09-252-991A-30052 (1-719)
QY 1235 AAGGCACTATGCTCTGCAACTTGTGGGGTTTACGGGAGCAGGAGTTCATGGGCT 1176
Dl 18 ArgGlyCysArgAlaCysArgAlaValProVal---AlaTrpSerGlyValGlySerAla 36
QY 1175 CTCCAGTAAGATGAGTCCAGGCTGCTGGATGTCCCAACCGTGGCTGAGG--- 1119
Dl 1175 ProPro---ArgGlyValLeuArgGlyProProAlaAlaLeuArgArg 50
QY 1118 -----GCGCAGGCTCAGCTATCTCTGATGCCATGTCCACGTA 1077
Dl 51 GlyAspArgGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 70
QY 1076 GCTGCTGAGTGGG-----GCTCCACTGCTGAGGCTGGGCTGCCAGAGG 1026
Dl 71 ArgProAlaAlaGlyArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 90
QY 1025 CCTGAAGCGCATGCTGTAGGCTTGGCTGAGAGATCATTTGGTGTGGGCTCCCTGACT 966
Dl 91 ProArgGlyAlaAlaAlaGlyVal-----ProAlaAla 101
QY 965 GGACACACAGACACATTGTGTAGGTCCTCTGAGGAATGACCTTGGTGGCAGGATCGGTG 906
Dl 102 SerArgAspArgArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 117
QY 905 TGTGAGAGCTGTCTCCGAGTGTGGCCAGGCGCAAGCGGTGGCCAGCTCATCTCTGGG 846
Dl 118 -----ProGlySerAsn----- 121
QY 845 TGATGGCGCGGCGCCAGCAGCTCCACTGTATCCCCAGGAGGCTGGCGGGAGCTGGAG 786
Dl 122 -----AlaGlyProSerGlnArgValValAlaAlaProGlySer----- 133
QY 785 TACTGTACAGGCTGTGGACCTGGTGTGGGTGAAGAGTCATCTCT---CATCATCTGAGA 729
```

```
DQ 134 ..... GlyAlaGThrArgProGlyArgGluArg 143
QY 728 GCGCTTCAACAGGAGGACAC...CTGGCATATCCGGATAGAGCTGGAGGCGATGCTCC 672
DB 144 CysLeuArgAlaProAlaArgPrValAlaGlyProProAlaGlyTTPArgGluLeuPro 163
QY 671 GGGAGAGAGAGTCACTCCAGGAGTGGAGCACTGCTGGTATAGAGAGTCCAGACACAGGA 612
DB 164 ProAlaGlySer...AlaArgAlaArgGly...ArgAlaProGluProGly 178
QY 611 CAATGGCATTCAGAGGGTGGHAGAGAGGAGGAGCAAGGATATCAAGCATATGGGATCAG 552
DB 179 Arg.....GluSerAlaGlyAlaGlyAlaGlyProThrProGlySerValArgAspArg 195
QY 551 CGAGACACAGAGAGAGTCTCTGGTGGAGAGAGGAGAGGAGAGAGAGAGAGAGAGAGAG 495
DB 196 ArgArgGlyArgProGlyArgCysProGlyAlaLeuAlaCysArgCysArgHisSerGly 215
QY 494 TGAGGCGCTGGGTGGCCACATGATGTCATCCAGAGAGACTCTCTATTGTCAGCATCTAA 435
DB 216 AspGlyAsnAlaArgAspValAlaGlyAspPleuValArgArgGlyTTPArgAlaAlaGly 235
QY 434 AGACGCGCT.....CCCTGTAGAGG 414
DB 236 ArgGlyProTyArgSerPheGlyLeuProLeuPheMetGlnValAlaThrLeuProArg 255
QY 413 AGCTGCTGTGGAGGCGAGTGTGCAGACCGAGACTCTC.....CCATGGGAG 366
DB 256 AlaLeuAlaAlaValLeuLeuLeuValGluGlyLeuLeuHisArgLeuPheArgMetGln 275
QY 365 CCACHT...TGTCACAGGTTCCATTTCGATGAGGTCAGCCAGGAGACTTTCGAGAA 309
DB 276 ProLeuGlnAlaPheGluGlnLeuAlaThrGlyLeuLeuGlyAsnAlaAlaGly 295
QY 308 CAGGACAGTGCAGCCAGTGTGATGCCATAGCCATAGAGTCAAGTGTGTGGTGTGTTT 255
DB 296 AlaGlyThrAspLeuAspLeuGluAlaValAlaMetGlnAspLeuHisLeuHisGly 315
QY 254 ...TTAGCTTCCGACACAGCAATATCA 231
DB 316 AlaGluLeuGlyGlnSerHisLeuGlnAlaAlaArgValAlaMetGlyYsAlaGlyAsp 315
QY 230 ...GATCAATCACTCAGAGTCTTACAGAG... 204
DB 336 LeuAspAlaProLeuHisGlyTTPGluGlnAlaArgGluLeuGlyGlnSerLeuPro 355
QY 203 ACTCTGGAGTTTCCAGAGAGATGAGTGTATTATATGATGATATCTATAGCCGCTAGCG 144
DB 356 AlaProAlaAlaGlySerGlnArgGlyGlnArgGlyGlnArgGlyGlnArgGlyGln 366
QY 143 AGATCTCTCCGATTCTGAGTGTGTTT...AGAGAGAA 106
DB 367 .....AlaValAlaGlyGlyProGlnGlyArgGlnHisLeuArgGlyGlnArgGly 382
QY 107 TAGACTTTGGAGTAAGTGGCTGA...TCAGCCA 78
DB 382 alValGlyGlyGlnCysGlyTTPLeGlyPheLeuGlyLeuThrHisLeuLeuSerAlaAla 402
QY 77 GCTTCACCCGAGTCCGAGTCTAGAGAGAAATCCCTTTCCTTTGGCTCTTC 24
DB 402 lalPheAlaAlaArgCysArgGlnAlaArgAlaLeuAspHisGlyLeuArgGlyAla 422
QY 23 TCTTTCTC...CTTGGAGT 9
DB 422 rgAspLeuGlyCysGlyPheArgGlnAlaGlyAlaGlyLeuTTPPro 438
```

RESULT 15

US-09-023 905A-2

: Sequence 2, Application: US/39023925A

: Patent No. 6473778

: GENERAL INFORMATION:

: APPLICANT: Roberts, Thomas M.

: APPLICANT: King, Frederick J.

APPLICANT: Harris, David F.
APPLICANT: Hu, Eiding
APPLICANT: Spiegelman, Bruce
APPLICANT: Chan, Connie
TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
FILE REFERENCE: EPN-021
CURRENT APPLICATION NUMBER: US/09/023, 905A
CURRENT FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/038,191
PRIOR FILING DATE: 1997-02-14
NUMBER OF SEQ ID NOS: 39
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2

LENGTH: 1129

TYPE: PRT

ORGANISM: Bos sp.

US-09-023-905A-2

Alignment Scores:

Pred. No.:	C.00121	Length:	1129
Score:	135.50	Matches:	109
Percent Similarity:	34.97%	Conservative:	55
Best Local Similarity:	23.24%	Mismatches:	163
Query Match:	5.89%	Indels:	142
DB:	4	Gaps:	22

US-09-735-251-3 (1-1273) x US-09-023-905A-2 (1-1129)

QY 62 CACTGGCGGTGAAGCTGGCTGACCGACCACTTACTCCAAAGTCTATTCTTCGGTTG--- 118

DB 609 HisLeuAlaValArgThrAlaAspGln-----ThrSerLeuHisLeuVal 623

QY 119 -----CCAGACAGAACTGGGAGAAATACG 145

DB 624 AspPheLeuAlaGlnAsnCysGlyAsnLeuAspLysGlnThrAlaLeuGlyAsnThrAla 643

QY 146 CTAGGAGGCTATAGTATTTCATTTCGAGCAGCTTATTCTGGCAAACTCCAGGAGTCT 205

DB 644 Leu...HisTyrCysSerMetTyrSerLysProGluCysLeuLysLeuLeuArgSer 662

QY 206 GTTCAGAGAGTGGAGTGGATTGAT-----CTGATCTACTGTGTCGGAAG 250

DB 663 LysProThrValAspValValAsnGlnAlaGlyGluThrAlaLeuAspIleAlaLysArg 682

QY 251 CTAAAGAGATACAGACACATTGACTTCTATGGCATTCACCTGGGTCCACT----- 301

DB 483 LeuLysAlaThrGlnCysGluAspLeuLeuSerGlnAlaLysSerGlyLysPheAsnPro 702

QY 302 ---GTTCATCTT-----CTGCAAAAGTCTTCGCT 328

DB 703 HisValHisValGluTyrGluTrpAsnLeuArgGlnGluMetAspGluSerAspAsp 722

QY 329 GAACCTGATAGAAACCGGACCTGTGCAGAAAGTGTGCTCCATGAGAGAGTTCGGGTG 388

DB 723 AspLeuaspLysProSerProIleLysLysGluArgSerProArgPro----- 739

QY 389 TTGCACATCGCTGCACAGCAGCTCTCTTACAGGAGGCGGTCTTTAAGATGCTCAGC 448

DB 740 ---GlnThrPheCysHisSerSer-----IleSer 749

QY 449 AATAGAGTGTCTGGATCAGATCATTTGTGGCACCCAGCCCTCAGCAGTACCCCTATT 508

DB 750 ProGlnaspLysLeu-----SerLeuProGlyPheSerThrPro----- 762

QY 509 GCTCTTGGGCTTCTCCAGGACAGGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 568

DB 763 ----ArgAspLysGlnArgLeuSerTyrGlyAlaPheThrAsnGlnIle 777

QY 569 ----GATAGCTTGTGCTGCTCACCACGCCCTCTCTCAATGCTGCTGCTGCTG 616

DB 778 PheValThrThrSerThrAspSerProThrSerPro----- 789

```
QY 617 GTTCTGCACTCGGTAGCAGGAGTCCCGCATGCTGGGACTGACTCTCTTCCCGAGC 676
DB      ||||| ||| |||||
QY 790 -----IleAlaGluAlaProLeuPro----- 797
DB      ||| |||||
QY 677 ATGCCTCCAGCTCATACGAGATATGCGAGTGGCTCTCTTGAAGGCTCTCAGAT 736
DB      ||| |||||
QY 798 ---ProArgAsnAlaThrLysGlyProSery----- 807
DB      ||| |||||
QY 717 GATGAGGAGTCTTACCCAAACAGCAGGCTTACAGCTCTACAGTAGTCCAGC--- 793
DB      ||| |||||
QY 908 ProProSerThrLeuProLeuSerThrGlnThrSerSerGlySerSerThrLeuSerLys 827
DB      ||| |||||
QY 794 -----TCCCGCCAGCTCCCTCGCTGGTACAGTGCAGCTGCTGTCGCCGCGCATCAGC 847
DB      ||| |||||
QY 828 LysArgSerProPro-ProProProGlyHisLysArgThrLeuSerAspProProSerPr 847
DB      ||| |||||
QY 848 CAGAGTGAAGTGGCCACCG----- 872
DB      ||| |||||
QY 847 OLeuProHisGlyProProCAsnLysGlyAlaValProTrpLysAsnAspValGlyProSe 867
DB      ||| |||||
QY 873 -----CCTGGCCAGAGTCTGGAGAGACAGCTTTCACA----- 905
DB      ||| |||||
QY 867 rSerSerLysThrThrAsnLysPheSluGlyLeuSerGlnInlSerSerThrGlySe 887
DB      ||| |||||
QY 906 ----CACCGACTCTCTGGACCCAGATCATCTCATCTGATGAGGACTTACCAATCTCTCTGGT 961
DB      ||| |||||
QY 887 rAlaLysThrAlaLeuValProAlaValLeuProLysLeuProGlnLysValAlaLeuAr 907
DB      ||| |||||
QY 962 GTCCAGTCAGGAGCGCCCATCAGCAATATCTTTTACCAAGTCTACAGCATSCGCTT 1021
DB      ||| |||||
QY 907 GlyThrGlu-ThrSerHis-----LeuSerLeuAspLysAlaAsnValProP 924
DB      ||| |||||
QY 1022 CAGGACTCTGGCAGCCGCAAGCTTTAGAGTCAATAGTATGCTTACCTGAGAGAGTACTAGT 1081
DB      ||| |||||
QY 924 roGluIlePhe-GlnLysSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 943
DB      ||| |||||
QY 1082 GACATGAGCATCCAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
DB      ||| |||||
QY 944 AspLeuProProLysProThrGluLeuAla-----ProLysProThrIleGly--- 959
DB      ||| |||||
QY 1142 TCCAGAGCCCTGGAGCTCATCTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
DB      ||| |||||
QY 960 -----AspLeuProProLysProGlyGluLeuProProLysPro 972
DB      ||| |||||
QY 1202 GAA-----CCCGAGCTAGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1249
DB      ||| |||||
QY 973 GlnLeuGlyAspLeuProProLysProGlnLeuAlaAspLeuProProLysProGlnVal 992
DB      ||| |||||
QY 1250 AAGGAGCTCCATCTTATCTT----- 1250
DB      ||| |||||
QY 993 LysAspLeuProProLysPro----- 1000
DB      ||| |||||
```

Search completed: November 12, 2003, 0:15:25
Tot time : 43.5 secs

GenCore version 5.1.6
Copyright (c) 1993 2003 CompuGen Ltd.

OM: protein - protein search, using sw mode

Run on: November 12, 2003, 06:19:33 : Search time 20 seconds
(without alignments)
1371.458 Million cell updates/sec

Title: US-09-735-251-4

Perfect score: 2133

Sequence: 1 MS:SCWHLAVKLADQPLPK..... (SMDYCHWEALMVRPSGV 416)

Scoring table: B:OSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 761*

1: PIR1*

2: PIR2*

3: PIR3*

4: PIR4*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	171	8.0	373	S54583	ubiquitin-like pro
2	164.5	7.7	536	B84547	probable ubiquitin
3	143.5	6.7	551	C84547	probable ubiquitin
4	141	6.6	528	247147	saatatic mucin icic
5	141	6.6	671	T24568	hypothetical prote
6	139.5	6.5	1093	T13530	AF13 protein - hu
7	138.5	6.5	154	T13443	yeast Bsk2 homolog
8	138	6.5	872	S10201	hypothetical prote
9	138	6.5	992	A31668	hypothetical prote
10	136	6.4	1468	A44343	nucleoporin - rat
11	135.5	6.4	1162	2JH0587	exo-alpha-sialidas
12	135.5	6.4	2715	T13049	eyelid - fruit fly
13	132.5	6.2	1952	T48814	hypothetical prote
14	130	6.1	1273	S58782	SEC31 protein - ye
15	130	6.1	1487	1ED8E1	immediate-early pr
16	127	6.0	398	S50507	excision repair pr
17	127	6.0	1487	1ED8E6	155k transcription
18	126.5	5.9	351	2S0754	hypothetical prote
19	126	5.9	565	T39863	zinc finger protei
20	126	5.9	1147	T42627	ADP-ribosylation f
21	125.5	5.9	622	T15467	hypothetical prote
22	125.5	5.9	815	B56706	extracellular sign
23	125.5	5.9	990	151618	nucleolar phospho
24	125	5.9	502	T18562	hypothetical prote
25	125	5.9	1613	S39059	protein BRG1 - hu
26	124.5	5.8	1676	A56558	anucleate primary
27	123.5	5.8	1744	A54973	tensin, cardiac mu
28	123.5	5.8	1792	A57075	tensin, chicken
29	123.5	5.8	3149	Q0865	BRP1 protein - hu

ALIGNMENTS

RESULT 1

S54583

ubiquitin-like protein DSK2 - yeast (Saccharomyces cerevisiae)

N/A: alternate names: protein YMR021.02; protein YMR276w

C: Species: Saccharomyces cerevisiae

C: Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 24-Sep-1999

C: Accession: S54583; S59344

R: Pearson, D.; Bowman, S.

submitted to the EMBL Data Library, May 1995

A: Reference number: S54582

A: Accession: S54583

A: Molecule type: DNA

A: Residues: 1-373 <PEA>

A: Cross-references: EXBL:249704; NID:G8255340; PIDN:CAA89774.1; PID:G8255342; MIPS:YMR2-

A: Experimental source: strain AB972

R: Higgins, S.; Rose, W.D

submitted to the EMBL Data Library, April 1995

A: Description: The yeast ubiquitin-like proteins are involved in spindle pole body dur

A: Reference number: S59343

A: Accession: S59344

A: Molecule type: DNA

A: Residues: 1-108, R' 110-295, R' 297-373 <BIG>

A: Cross-references: EMBL:L40587; NID:G786149; PIDN:AAB07267.1; PID:G786151

C: Genetics:

A: Gene: SGD:DSK2

A: Cross-references: SGD:S0004889; MIPS:YMR276w

A: Map position: 13p

C: Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

F: 3-77: Domain: ubiquitin homology <UBH>

Query Match 8.0% Score 171; DB 2; Length 373;

Best Local Similarity 22.4% Pred. No. 0.0011;

Matches 83; Conservative 60; Mismatches 127; Indels 100; Gaps 13;

QY 63 LIYCGRKLKDDOTLDFYGIQPGSTVHVLRKSWPEPQKP----- 101

DB 44 LIYSKILKDDCVTSYHIOGHSVHLV-KSQPKPTASAAAGANNATATGAAGTGATFN 102

QY 102 -----EPVDKVAAMREFVLHT-----ALHSSSSSYREAVFK----- 132

DB 103 MSSGGSAGNPLADITSARYACYLNPADMPGPGGALNNDSNNQDELLRMNENPFGS 162

QY 133 ----MLSNKESLQOITVATPGLSS-DPIALGVLOKOLFVSVPADPNMLDTLVPAHPALVN 187

DB 163 QNVEKLNPNQLDFWIOQNPOIQMGPOARQMLQSPMFRQMLTFPDM-----IR 211

QY 189 AIVLVHVSAGAPMPGPTDSSSRSPSSSYRMPGGLFEGLSDEDDDFHPTNTSTPSS 247

DB 212 QMQFARMCDPNAGMGSAGGAASAPPA-----PG-----DAPEGSNTNTTSSN 257

QY 248 TPSSRPASLCYSGAAGPRPIQOSEALATALALASTPESSHTPTPTQCHSGSGTSPMSGV 307

DB 248 TPSSRPASLCYSGAAGPRPIQOSEALATALALASTPESSHTPTPTQCHSGSGTSPMSGV 307

253 QAVESDREAVKDEVIV..... .. PLVER "SNQKLMQVLTST"..... .. 289

156 ALGVQDKQLFSVFADPNMLDTVPAPALVNAIVLVVHSVAGSAAPVPGTD--SSSRSM 212

A:Note: sequence extracted from NCBI backbone (NCBI:U79200, NCBI:U79212)
C:Superfamily: human herpesvirus 4 nuclear antigen EBNA-3C

Query Match: 6.5% Score 139; DB 2; Length 992;
Best Local Similarity 22.5%; Pred No. 0.5;
Matches 71; Conservative 35; Mismatches 116; Indels 94; Gaps 15;

QY 177 TLVPAHPALVNAIVLVLS --VAGSAPYCT 255
DB 461 TVERAHPTEVPMVILHCPVPEVPPVPPKPTPTPSRPPGACVYDDEVLVVDVETTE 520
QY 206 DSSSSMSSSYRDVPGGFEGLSD... DEDEPHENTRTSSATPPSPPLASG 257
DB 521 DSSSVCPKPKRHKQDQFQSGREQKRAATVSPSTGPFAPVPAAGPAAAGPAA 580
QY 258 YSG----RAGRPITQSELATALASTPSSSHPTFGTQ..... 294
DB 581 PPAAGPAAAGP-PAAGRPILAPLS-AGPAAAGPILVTPPSARPRVAPPVVRMPVRQL 638
QY 295 GHSSGTSMS--GVCSGTPTINDLESQALCHALQASGQSL-QSCWQPLQCL--RQWG- 349
DB 639 POSTGRKQCFMEMRAGREITQ--MQEPSSHLQSAQTPTPRSWAPSYCALSUMDAGK 695
QY 350 ----ICDELSRPPCPVGTGSKPSSSLMEPEH... --EUPASPPFASCRGY 395
DB 697 AQPIESSHLSMSPTQPSHEEQP- EYEDERAPLDSLRHVAQAQAPAPVQGYQE 752
QY 396 -----CPKEALMKVPP 406
DB 753 PPAPCPVQGYQEP 466

RESULT 10
A44345
nucleoporin . rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 10-Jun-1993 #sequence_revision 14 Nov 1994 #text_change 63 Mar-1995
C:Accession: A44345
R:Sukenaga, J.; Blobel, G.
C: 72, 29-18, 1993
A:Title: A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and
A:Reference number: A44345; XMD:19333735; PMID:8422679
A:Accession: A44345
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-1468 <8K>
A:Experimental source: liver
A:Note: sequence extracted from NCBI backbone A:BIN:12145, NCBI:P123338;

Query Match: 6.4% Score 139; DB 2; Length 1468;
Best Local Similarity 23.4%; Pred No. 0.5;
Matches 90; Conservative 45; Mismatches 116; Indels 119; Gaps 17;

QY 31 GYSILG---GYSISLKQLIAGKQESVPGTELLGAGYQKRL --KCO 73
DB 1073 GGFTEGKVDASALSPSMFVLQRTKQCEPVTSLVEFKKALQNEBPXCVFSPENSE 1132
QY 74 QTLDFYGIQGSTVHLVLRKSNPEQKPE..... --PVQVAMREFRVHTALHS 122
DB 1133 QTKDESSSKPTSFSAKPSVKESDCLAKAFAPNQNTTTCQAKAPAFSLNSSSS 1192
QY 123 SS-----SYPEAVFKMLNKESLDLIIVATPLSSSPIALGVQKLFSPVADPNMLTL 178
DB 1193 SSTPATSSASIFGS-STSSSSFPVAAFFVQASNPVS..... --SSAFNSAESSTIS 1242
QY 175 VP-----AHPALVNAIVLVLSVAGSAP --VGTSSSSSYRPSVSPVPMQPGFEEGLS 230
DB 1243 QPLIFFQDCKPATTS..... --STASAPFVFGQASNSSTVSS --GTFEGATT 1298
QY 231 DEDDFHPNTRTPSS..... --STFSSPPA SLOYSMAAGPPPTQSELA 274
DB 1289 -----TSSSSGSFFVFGTGHASAPASPAFGNQFTTFQOSAGAQPPPSFGSISS 1339

QY 275 ALALAS-----TPSSSHPTPT-----PGTQGHSSGTSMSVSGVSGTPTINDLP 318
DB 1340 STALFSAGSQPVPPPTPTGTVSSSQPPVFGQQPSQSAFGSGTANASSVFGSGSTTNFNF 1399
QY 319 SQ-----A-QHALQASQPS 333
DB 1400 TNNPSPGVFTGASPSPTPAAAAQPS 1424

RESULT 11

JH0557
exo-alpha-stalidase (EC 3.2.1.18) - Trypanosoma cruzi
N:Alternate names: neuraminidase
C:Species: Trypanosoma cruzi
C>Date: 10-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-Jan-2000
C:Accession: JH0557
R:Perreira, M.E.A.; Meira, J.S.; Ortega-Barria, E.; Matzilevich, D.; Prioli, R.P.
J. Exp. Med. 174, 179-191, 1991
A:Title: The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial ne
A:Reference number: JH0557; MUID:91277609; PMID:1711561
A:Accession: JH0557
A:Molecule type: DNA
A:Residues: 1-1162 <8K>
A:Cross-references: GB:X61732; NID:9162302; PID:9162303
A:Note: the authors translated the codon TCT for residue 45 as Cys
C:Comment: This protein plays a role in parasite-host cell interaction.
C:Superfamily: trypanosomatid-specific surface antigen
C:Keywords: glycoprotein; glycosidase; hydrolase
F:394/Binding site: carbohydrate (Asn) (covalent); #status predicted

Query Match: 6.4% Score 135.5; DB 2; Length 1162;
Best Local Similarity 21.8%; Pred No. 0.9;
Matches 98; Conservative 51; Mismatches 189; Indels 111; Gaps 18;

QY 6 WELAVKLACQLPTPKSILRLPETELGEYSJGYSISFLKQLIAGKQESVDPDELID--- 62
DB 499 YREVLTMAN-----KIGSVYIDG-----ELLEGGQTVPDGTPDISH 537
QY 63 -----LIYQCKLKDQTLDFYGIQPT--GSTVHVLRKSWPEPD 98
DB 538 FVVGKYSQDVTISHV--VNNVLLNRRQLNTEIRTLFLSODLIGTEAHWDSSDSSAH 597
QY 99 QKPE-PVQKVAAMREFRVHLTHLSYSSRYEAVFKMLNKESLDLIIVATPLGSSDIAL 157
DB 598 STPTPADSSSAHSTPTPDSSSAHSTPTPADSSAHGTPSTPDSSAHGTPSTPADSSAH 657
QY 158 GYLQDKLQFVFPADPNM-LDTLVPAHPALVNAIVLVLSVAGSAPMGTQSSSRMPS-- 214
DB 658 GTPSTPVSSSAHSTPTPDSSSAHSTPS-----TPVDSSAHGAPSTPADSSAHGTPSTP 711
QY 215 --SSYRLVPGGFPEGLSCDEDDFHFNTRSTPSSRPSALGYSGAAGPRITQSEL 272
DB 712 VDSSAHGTP-----STPADSSAHSTPTPADSSAHSTPTPADSSAHSTPTPADSSAH 761
QY 273 ATALALASTPSSSHPTPTGT-----QGHSSGTSMSVSGVSG--GTPITNDLFSQALQHAL 326
DB 762 SAAHGTPTPADSSAHSTPTPADSSAHGTPSTPADSSAHGTPSTPADSSAHSTPTPDSS 810
QY 327 QASGQPSQSQWQPLQQLRWGIDDELRLRCP-----PVGTSKOPWSSSLLEPHEL 382
DB 811 SAHGTPTPT-----VDSSAHSTPTPDSSAHGTPSTPDSSAHGTPSTPDSS-----AHST 851
QY 383 PASP-EPASCSGYPWEALMKVPSLPV 410
DB 852 PSTPADSSAHSTPTPADSSAHGTPSTPV 880

RESULT 12

T11049
eyelid - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T11049

R:Treisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z17592
A:Accession: T13049
A:Status: preliminary; translated from CE/EMBL/GenBank
A:Molecule type: mRNA
A:Residues: 1-2715 <TR>
A:Cross-references: EMBL:AF053091; NID:G2981220; PID:G2981221; PDCN:AA06254.1
C:Genetics:
A:Gene: eld
A:Cross-references: Flybase:FBgn:0000313
C:Function:
A:Description: could act as a transcription factor antagonistic to the wg pathway
C:Keywords: DNA binding

Query Match 6.4% Score 135.5; DE 2; Length 2715;
Best Local Similarity 22.4%; Pred. NC: 2.8;
Matches 93; Conservative 41; Mismatches 143; Indels 133; Gaps 19;

QY 93 SWPEPDQKPEPVOKVAMRE-FRVJHCTALESSSYPEAVEKYLSEKES-LDOQ-----143
DB 956 NWPEPAGSPQVNVHVPVCEPPPTITTKKSCSLCKLYEMCTUPDRRWLDKLRAYE1014
QY 144 -----IVATPGLSDPIAL-----WVQXK-----CLPSEVADPMYLDL178
DB 1015 ERRTPTACPTIGKQPLDLYLVYVYKROGFEVTKSKTKAFAGLIGIGASSAAAYTL1074
QY 179 VPAHPALVAIVLVLSVAGSA-PVY-----QTSSSSMSSSVKRCMPCGGFLPEGLSDQ232
DB 1075 RKHYTK-NLLTTECHFRDQIDPIILQVVAASKKTKAKAATPSPGSEHLNG---1128
QY 233 EDDHPNTRTPSSSTPSRFALEQ-----YSSAAGRPITQSELAT274
DB 1129 -----TTNGTSGNSQSEPPAPAEAPAAALVYGVYGGSPYFGASQCP--DYAT1178
QY 275 ALALASTP-ESSSHITPPTGQ-----GHSSITSEF-----302
DB 1179 AGVQVRPSONNCTPHPGMAAAVAGCNVKNPNFEPPIAAGTGPSTGPGQGGFGP1238
QY 303 -----MSSGVQSGTPTTNLFQALHQAQSGQ-----PELQSQWQPGIQQ344
DB 1239 GAARGGAGAVGAGVGPGPHPPPHFPHFAAGCAAGQCCCHFHQHPGLPSPPPQQQQ1298
QY 345 LRDMGIGQDELRLRPVPPVGTGSKQWSSLLFEPHLPASPEP---PASCRCGYCP397
DB 1299 ---GQGGQQ-----PPRVSQVIRPPAPQSLHQQQFELHJVVVFAAAAPPP1341

RESULT 13
T48814
Hypothetical protein J556.220 [Imported] Neurospora crassa
C:Species: Neurospora crassa
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C:Accession: T48814
R:Schulte, U.; Aign, V.; Hohelsel, C.; Biardi, P.; Faltus, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224541
A:Accession: T48814
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1952 <SCH>
A:Cross-references: EMBL:ALJ53822; GSPDE GNC0112; NCSP:J556.220
C:Experimental source: cosmid contig J554; strain 74
C:Genetics:
A:Gene: NCSP:J556.220
A:Map position: 2
A:Introns: 281/3

Query Match 6.2% Score 132.5; DE 2; Length 1952;
Best Local Similarity 22.4%; Pred. NC: 2.8;
Matches 79; Conservative 45; Mismatches 140; Indels 99; Gaps 16;
QY 81 TQPGSTVHLRKSFWEPDQKP-EIVQKVA-MREERVH-ALHSS-SYSEAVFQMLSN136

DB 1652 VTPAPAAHTACQVAQPVAPVPTQSPVQSVACHVTSSQASTTAAHPVA-----1699
QY 137 KESLDIOIVATPGLSSDPIALGVLDQKDLFSVFADPNMLDTLVPAHPALVAIVLVLSV196
DB 1700 -QSVPRPVNSNPTSAAPVAAGCTAQ-----VASATV-----TQPHRALSSVSQSL1746
QY 197 AGSAMPCTGSSSSSMPSRSSYRMPGGLFEGLSDDDDHFNTRSTPSSS--TPSSR--252
DB 1747 PQSVP-HAAQCAAHQTPHSASRPV-----QSVPOSVPQATPQAVPRSTGALTPTAOPG1800
QY 253 PASLGYSSAAGRPITQSELATALALASTPSSSHPTPTGTQCHSSGTSMPSSGVQSGCTP312
DB 1801 PVSFAVSGSGVPAP-----SAAQSVAPAPVSS--TPVP-----AATVAPASTVAAATTP1847
QY 313 TNDLFSQALQHAL-----QASQPSLQSQWQPOLQQLRDMGIGQDELRLPCR361
DB 1848 TRVTAAPALSAATNPAPVPSPQPHGCTGQAPAAQORPPAQO-----APAT1894
QY 362 PVGTGSKQPKSSSL-----EEPHLPASPEPPASCRCGYCPWEALMKVPPS407
DB 1895 PTITSAAPPPPTLAPPPTPPPPPTEDRPPPPPP-----PAEAPPPPPPT1941

RESULT 14

S58782
SEC31 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein D1299; protein YDL195W; WEB1 protein
C:Species: Saccharomyces cerevisiae
C:Date: 29-Nov-1995 #sequence_revision 24-May-1996 #text_change 26-May-2000
C:Accession: S58782; S48241; S67750; S72065
R:Verhasselt, P.; Voet, M.; Volckaert, G.
submitted to the EMBL Data Library, December 1994
A:Reference number: S58777
A:Accession: S58782
A:Molecule type: DNA
A:Residues: 1-1273 <VER>
A:Cross-references: EMBL:X83276; NID:G1004294; PID:G1004300
R:Zierer, H.A.; Walberg, M.; Berg, P.
submitted to the EMBL Data Library, September 1994
A:Description: Suppression of mutations in two S. cerevisiae genes by the adenovirus
A:Reference number: S48241
A:Accession: S48241
A:Molecule type: DNA
A:Residues: 1-316 'T', 319-366 'T', 368-690 'N', 692-753 'V', 755-876 'A', 878-1273 <ZIE>
A:Cross-references: EMBL:J15219; NID:G557057; PID:G557058
R:Volckaert, G.; Verhasselt, P.; Voet, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67745
A:Accession: S67750
A:Molecule type: DNA
A:Residues: 1-1273 <VOL>
A:Cross-references: EMBL:Z74243; NID:G1431319; PID:e253103; MIPS:YDL195
A:Experimental source: strain S289C
R:Verhasselt, P.; Voet, M.; Mathys, J.; Volckaert, G.
Yeast 12, 1065-1070, 1996
A:Title: The sequence of 23 kb surrounding the SNF3 locus on the left arm of yeast ch
ing putative genes for ribosomal protein L35 and a sugar transport protein.
A:Reference number: S72060; MUID:97051595; PMID:8896272
A:Accession: S72065
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1273 <VEW>
A:Cross-references: EMBL:X83276; NID:G1004294; PID:G1004300
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 199.
C:Genetics:
A:Gene: SGD:SEC31; WEB1
A:Cross-references: MIPS:YDL195W; SGD:S0002354
A:Map position: 4L
C:Superfamily: unassigned WD repeat proteins; WD repeat homology
F:253-287/Domain: WD repeat homology <WPI>

Query Match 6.1% Score 130; DB 2; Length 1273;

```

Best Local Similarity 22.38; Pred. No. 2.1;
Matches 79; Conservative 50; Mismatches 160; Indels 56; Gaps 16;
QY 78 PYGQPGSTV--HVLKRSNPEPKVQKVPKVAAMPEFRVLHTALHSSSYREAVFKVL 135
Db 8/8 PAPIQPMPIGMDYNAQSSIPSQ PPI NAVES---QTPHLNPRANDQKDLPUKKE 931
QY 136 NKESLDQIIIVATPOLSSDRIALGVLOQKLFVFAAPKNVLDLTVAFPAIYVA--IVLV 192
Db 932 KPSRAKAVUAPPNLTPTPLNGI--PANAATVF--PPISTAFSSVSV 979
QY 193 ----LH-----SVAGAPVPGTDS--SPNPKSNVFCVNGGKIFGLSDECECHPHNTE 242
Db 960 SPPELHXNSRVSLVATSESPRASNPNYAPPCSSQCFRIGI STANGTNTAQVASE 1037
QY 243 TPSSSTSSSPASIGYSSAGAPPHITUSCLATAALAEFTESSEHPT LPTCHHS--SPT 100
Db 1038 NPVAPPCCQVAT--PLSGVFFFAFLKASNPYATATTONHSSSYPTVEYNNHMTSF 1096
QY 301 SPMSGVCSGP-----ITNDFEALCHALCASQPSLQSKWLPQLQLLDVGIQDE 154
Db 1097 PPVFNKEPTGPPPSMKKRNKLAS--FCNPSSQATVPFLLSSSNAP----- 1142
QY 355 LSLRPPRPVGTSKQWSSLLSEFHELPAPEPPASCRGVCYCHWEALMVFESLP 469
Db 1143 --LQFSQPTLASGV--NTSAENVGEEFADQQAIVQELK--BELARVTELP 1189

RESULT 15
EDBEE:
Immediate-early protein - equine herpesvirus 1
C:Species: equine herpesvirus 1
C>Date: 30-Jun-1990 #sequence_revision 10 Jun-1990 #ext_change 16 Jul-1999
C:Accession: A33764
R:Grundy, F.J.; Baumann, R.P.; O'Callaghan, D.C.
Virology 172, 223-236, 1989
A:Title: DNA sequence and comparative analysis of the equine herpesvirus type 1 immediate-early protein
A:Reference number: A33764; MJD:097034; PMID: 2549711
A:Molecule type: DNA
A:Residues: 1-1487 <GRU>
A:Cross-references: GB:04166; NID:031010; PID:0000082; PID:010911
C:Superfamily: herpesvirus immediate-early protein 12.19
C:Keywords: DNA binding; early protein; transcription regulation

Query Match 6.18; Score 130; EPI 1; Length 1457;
Best Local Similarity 21.24; Pred. No. 2.8
Matches 85; Conservative 43; Mismatches 116; Indels 136; Gaps 16;
QY 48 AKKJGESVPEDELINQGHKVI QTLVAV--LPTVPP EFWQWPK 100
Db 30 ASAAEEGFTQFAAPDQVAVG--VAV--PQVLESTLWVPPRHEPEFGGKA 81
QY 101 -----PPVKKVAMSEERVLTALHSSSYREAVFKVL 135
Db 82 KRSPSAAGSGPPTFAAAQFASPAFSPASGLAM--LKWHSNAPGNRP----- 131
QY 136 NKESLDQIIIVAT-----PGLSDFIALGVLOQKLFVFAAPKNVLDLTVAFPAIYVAI 189
Db 132 -----ATGSSPGGGDAADPAVAL-----DSCIETTPGSPQ-----PEPFSSA--- 168
QY 190 VLVLHVSAGAPMPTDSSSRSPSSSYRMPQDFLEGL-----SQDEIDFHNT 240
Db 169 -----SPGGSPAPRVRSISISSSS--SSSMGLDQADGAGASSSSSSSSSSSSSSSSSSSS 223
QY 241 RSTP---SSSTPSPRPASLGYSG--AAGPRPTGSELATACALASTPES--SHTPTPTGTQ 294
Db 224 EETPRPHSQNAKTTPSAAGSPGSSGGDRP-----AAGAA--PKSCPSAASPGAP 274
QY 295 GHSSGTSPMSGVSGSTFTINDLFSQALQHALASGQPSLQSKWLPQLQLRONGIQDDE 354
Db 275 APAPASAPAPSPGGLLPFGARILEYLEGVREANLAKTLE----- 315

```

```

QY 355 LSLRPPRPVGTSKQWSSLLSEFHELPAPEPPASCRG 394
Db 316 ---RP-EPPAGVASPPGSR-----PHRLPKDQRPKSALAG 346

```

Search completed: November 12, 2003, 06:39:07
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 06:20:53 : Search time in seconds

Without alignment:
1334173 Million cell updates/sec

Title: US-09 735 251 4

Perfect score: 2133

Sequence: 1 NLSLSWH:AVKLAQCFUPK... : SWISSPROT:MKVSLPVP 41

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47526705 residues

Total number of hits satisfying chosen parameters: 121463

Minimum DB seq length: 0

Maximum DB seq length: 203000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_41*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	170	8.0	373	1	DSK2_YEAST
2	143.5	6.7	555	1	DAB2_HUMAN
3	141.5	5.6	555	1	Q8GKA6_maca_fasc
4	139.5	6.5	1093	1	AT17_HUMAN
5	138.5	6.5	354	1	YAC9_SCHPO
6	138	6.5	992	1	EPN6_EBV
7	136	6.4	468	1	N553_RAT
8	135.5	6.4	462	1	TYNA_EPRYE
9	133.5	6.3	907	1	AF3_HUMAN
10	130	6.1	1273	1	W811_YEAST
11	130	6.1	1457	1	1004_HUMAN
12	129	6.0	598	1	NX41_CACAE
13	127	6.0	398	1	R421_YEAST
14	127	6.0	1487	1	10F4_HSVB
15	126	5.9	565	1	SRL1_SCHPO
16	125.5	5.9	815	1	MC27_HUMAN
17	125.5	5.9	324	1	1K52_HUMAN
18	125	5.9	676	1	MUC1_MEJAL
19	125	5.9	707	1	BC16_MCUSE
20	125	5.9	5120	1	PCO2_HUMAN
21	124.5	5.8	1676	1	APSA_EMENT
22	124	5.8	1233	1	HEN4_HUMAN
23	123.5	5.8	1744	1	TENS_CHICK
24	123.5	5.8	3149	1	TECU_EBV
25	123	5.8	1844	1	PCUR_TYVYC
26	122.5	5.7	1210	1	AF4_HUMAN
27	122	5.7	1305	1	GAK_RAT
28	122	5.7	1487	1	MD53_YEAST
29	122	5.7	1844	1	PCUR_TYVY
30	121.5	5.7	706	1	BC16_HUMAN
31	121.5	5.7	3421	1	TECU_HSVB
32	121	5.7	2161	1	SHK1_HUMAN
33	121	5.7	3530	1	MY15_HUMAN

34	120.5	5.6	1465	1	NCO2_RAT
35	120.5	5.6	5147	1	PCO2_HUMAN
36	120	5.6	770	1	DAB2_HUMAN
37	120	5.6	1983	1	TF20_MOUSE
38	119	5.6	1367	1	AYH1_YEAST
39	119	5.6	1794	1	YDC9_SCHPO
40	118.5	5.6	1206	1	P3PA_HUMAN
41	118.5	5.6	1122	1	GAK_HUMAN
42	118	5.5	269	1	NO20_MEJTR
43	118	5.5	145	1	T0B1_HUMAN
44	118	5.5	971	1	RJHC_HUMAN
45	118	5.5	2404	1	SON_MOUSE

ALIGNMENTS

RESULT 1	DSK2_YEAST	STANDARD	PRT: 373 AA
ID	DSK2_YEAST		
AC	P48510		
DT	01-FEB-1996 (Rel. 33, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Ubiquitin-like protein DSK2		
GN	DSK2 OR SHE4 OR YMR276W OR YMR021.02		
OS	Saccharomyces cerevisiae (Baker's Yeast)		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
CX	NCBI_TaxId:4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=S288c / AS472;		
RC	PubMed=9169872;		
RA	Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T., Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S., Jagels K., Lyne S., Moule S., Odel C., Pearson D., Rajandream M.A., Rice P., Skelton C., Walsh S., Whitehead S., Bartell B.G., Xie M., et al.		
RT	"The nucleotide sequence of Saccharomyces cerevisiae chromosome XII."		
RT	Nature 387:90-93 (1991).		
CC	FUNCTION: INVOLVED, WITH RAD23 IN SPINDLE POLE BODY DUPLICATION.		
CC	SUBCELLULAR LOCATION: Nuclear (Probable).		
CC	SIMILARITY: TO S.POMBE SPAC26A3.16.		
CC	SIMILARITY: Contains 1 ubiquitin-like domain.		
CC	SIMILARITY: Contains 1 UBA domain.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).		
CC	EMBL; L40587; AAC37267.1; ..		
DR	EMBL; Z49704; CAAB9774.1; ..		
DR	PIR; S54583; S54583.		
DR	HSSP; Q15843; 1ND8.		
DR	SGD; S0004989; DSK2.		
DR	InterPro: IPR006636; ST1.		
DR	InterPro: IPR000449; UBA_domain.		
DR	InterPro: IPR000626; Ubiquitin.		
DR	Pfam; PF00627; UBA.		
DR	Pfam; PF00240; ubiquitin; 1.		


```

RESULT 3
DAB1_MAFCA STANDARD; PRT: 555 AA.
AC Q9BGX5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Disabled homolog 1.
GN DAB1.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecidae; Macaca.
CX NCBI_TaxID=9541;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Frontal cortex;
RA Csada N., Hida M., Kusuda J., Tanura K., Isaki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.
RT "Isolation of full length cDNA clones from macaque brain cDNA
RT libraries."
RL Submitted (FEB-2001) to the EMBL/GenBank/CCDC databases.
CC -!- FUNCTION: Adapter molecule functioning in neural development. By
CC similarity.
CC -!- SUBUNIT: Associates with the SH2 domains of Src, Fyn and Abl. By
CC similarity.
CC -!- DOMAIN: THE PTD DOMAIN SPECIFICALLY BINDS TO THE ASN-PRO-XAA-
CC TYR(P) MOTIF FOUND IN MANY TYROSINE-PHOSPHORYLATED PROTEINS.
CC -!- PTM: Phosphorylated on Tyr-198 and Tyr-220 upon insulin induction
CC in embryonic neurons (By similarity). Also phosphorylated on Ser-
CC 491 independently of insulin signaling (By similarity).
CC -!- SIMILARITY: Contains 1 PTD domain.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB055282; BAB21906.1;
CC InterPro; IPR006C23; PTD_PTD.
CC Pfam; PF06440; PTD_1.
CC SMART; SM00462; PTD; 1.
CC PROSITE; PS01179; PTD; 1.
CC Developmental protein; Neurogenesis; Phosphorylation
CC DOMAIN 36 189
CC MCC_RES 198 198 PROSEREYKATID BY SIMILARITY
CC MOD_RES 220 220 PROSEREYKATID BY SIMILARITY
CC FT MOD_RES 491 491 PROSEREYKATID BY OVER 1BY
CC FT SIMILARITY
CC SQ SEQUENCE 555 AA; 59920 MW; A3B02A1D3C4B51 82544.

Query Match 6.6%; Score 141.50; E-Value 1.0; Length 555
Best local Similarity 21.8%; Pred. No. Gaps 18;
Matches 83; Conservative 52; Mismatches 150; Indels 95; Gaps 18.

QY 67 GRKLKDDQTLDFYGIQPGSTVHVLRKSWDEPKPEPV...DKVAMSEFRVL 116
DB 57 GDLKCDMMKLGKVAGA...RKSEHKQKFLTISFGIKLFDKTCALQHHA 110
QY 117 HTALH-SSSSYRAVPMK-SNESLDCQIVATPG-SSODIALGV-CKDLSVFA 175
DB 111 HEISY-ASYKTDHRAFYGACGKGNHFRVAIKTAAGSEV...LCLRLFL 167
QY 176 DTL-----VPAHPLVNAIV-----LVLSVAGSAPV--PQTS 215
DB 169 EELEKKAQKQCEQAVYQTI-EEEDVEDPVYQVYFE--AGHEFINDP 225
QY 216 SYRDMPGFLFEGLSDEDDFHNTSTPSSSPSRPASLQYSGAAGPRPT 275

RESULT 4
AF17_HUMAN STANDARD; PRT: 1093 AA.
AC P55198;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE AF-17 protein.
GN MLLT6 OR AF17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=94336695; PubMed=8058765;
RA Prasad R., Leshkowitz D., Gu Y., Alder H., Nakamura T., Saito H.,
RA Huebner K., Berger R., Croce C.M., Canaan E.
RT "Leucine zipper dimerization motif encoded by the AF17 gene fused to
RT A2-1 (MLL) in acute leukemia."
RJ Proc Natl Acad Sci U.S.A. 91:8107-8111 (1994).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL
CC TRANSLOCATION T(11;17)(Q23;Q21) THAT INVOLVES MLLT6 AND MLL/HRX.
CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.
CC -!- SIMILARITY: Contains 2 PHD-type zinc fingers.
CC -!- SIMILARITY: HIGH TO AF10.
CC -!- DATABASE: NAME=Acias Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/Chromocancer/Genes/AF17.html".
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U07932; AAA21145.1;
CC PIR; I38533; I38531
CC Genew; HGNC:7138; MLLT6.
CC MIM; 60328;
CC GO; GO:0007048; P oncogenesis; TAS.
CC GO; GO:0006355; P regulation of transcription, DNA-dependent; TAS.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF06628; PHD; 1.
CC SMART; SM00249; PHD; 2.
CC PROSITE; PS01359; ZF_PHD_1; 1.
CC PROSITE; PS00016; ZF_PHD_2; 2.
CC Nuclear protein; Zinc-finger; Repeat; Chromosomal translocation;
CC Proto-oncogene.
CC ZN_FING 5 57 PHD-TYPE 1.
CC ZN_FING 51 68 C4-TYPE.
CC ZN_FING 111 180 PHD-TYPE 2.
CC DOMAIN 190 211 GLY/SER-RICH.
CC DOMAIN 275 282 POLY-SER.
CC DOMAIN 326 338 POLY-SER.
CC DOMAIN 729 764 LEUCINE-ZIPPER.

```



```

FT DOMAIN 822 829 POLY-SER.
FT DOMAIN 834 862 PRO-RICH.
FT DOMAIN 935 984 GLN-RICH.
FT DOMAIN 1040 1051 POLY-ALA.
FT DOMAIN 1069 1080 GLY-RICH.
FT SITE 551 551
SQ SEQUENCE 1093 AA, 112021 MW, 175342.60 kDa, 553E CRC64.

Query Match
Best Local Similarity 21.4%; Pred. No. 0.5;
Matches 108; Conservative 48; Mismatches 161; Indels 161; Gaps 30;

QY 1: KLAQQT---PKSLPLPTE/GEVSLGYS:SLKGLIAG- - - - KLCSSVP 56
Db 590 RLRSPTSLPSSASISTQV---SLACSTFSPSTHPTGPMGAVNPLSGAERSHT 647
QY 57 DPEDLHYCKRLKDDQTLDFYGIQNGTVV - - - - -LRKSWPEPDQKEE 102
Db 648 EPDLEDCSPCRGTSPOESLS- - - - -SMGPIASLPALPDOTASPCGGQLDPAAPGTINWQ 705
QY 103 PVDK- - - - -VAAMREPRVLHLSHSSSYREAFKMLSNKES:DCIIVATPGLSDDP: 155
Db 706 LSEKGGGCGEAGNIVEMKADRLKQENRLOQWLSLCTAKKERQILNV - - - - 755
QY 136 ALGVLOOKLFSVFAELPM:ETLVPAHVALVNAIVLHLSVAGSAPMPTD- -SSRSKM 212
Db 756 - - - - -QLSVPEFALPAALPAAGPPVPGYGLPQALSSSS-LSKSP 797
QY 213 PSSSYRMPGQFLPEI:SDDEDFHNTSTSSRPSASLYSQAAPRPIITSEL 272
Db 798 POKSSIG - - - - -UNSLSTSEQ - - - - -RIGCFPSRSSSSSEHFTIPLQQSE 844
QY 273 AT-ALAIAGTPESSHTF- - - - -TPTQGHESSTSPSSGYGOSTPTPTPLFSCA:QHA 325
Db 845 ATTLALPGAPAP:PPQPONGLGAFAPAG - - - - -LAVEMAEGLGGLAGNGLPUNGLLG 902
QY 326 L - - - - -CASQPSLO - - - - -SCNQLQCG:RPMHLODDELSL 357
Db 903 LNGAAANFASLSAGGAPTCLQDPMNSLTELCEHFLQCGGQQGQ - - - - -QLLA 955
QY 358 RPKAPP - - - - -VATSKPNSSESLSE- - - - -PHELPA 394
FT 956 SPQTPERTVYQMIQI:CKKRELQCLQVAG:SLCPMASLIAGSSTLMSAS:PGLEPT 1015
QY 395 SPEPPANCRCVCPWEALMAYVEPSL 498
Db 1016 ASATP- - - - -LAFANAL VAPG- - - - -

```

RESULT 5

```

YALG SCHPO
ID YALG SCHPO STANDARD; PRT. 354 AA.
AC Q12169.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DI 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein C26A3.16 in chromosome 1.
GN SPAC26A3.16.
OS Schizosaccharomyces pombe (Fission Yeast)
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes.
CC Schizosaccharomycetales; Schizosaccharomycetaceae.
CC Schizosaccharomycetes.
OX NCBI_TAXID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21849401; PubMed=11859360;
RA Wood V., Williams R., Rajandream V.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
SA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell D., Fraser A.,
RA Goble A., Hamlin N., Harris E., Hider G., Hider G.,

```

```

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jageis K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
RA Rutherford K., Putter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds K., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voickacit G., Aert R., Robben J., Grymonprez B.,
RA Weijens L., Vansteens E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer T., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Putneller B.,
RA Goffeau A., Cadieu A., Dreano S., Gloux S., Lelaune V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardet C., Tallada V.A., Garzon A., Thode G.,
RA Dominguez A., Revuelta J.L., Moreno S., Armatrong J., Forsburg S.L.,
RA Certutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Messy D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-890(2002)
CC - - - SIMILARITY: TO YEAST DSK2
CC - - - SIMILARITY: Contains 1 ubiquitin-like domain.
CC - - - SIMILARITY: Contains 1 UBA domain.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL: 269240; GAZ9239.1;
CC PIR: T33404; T38404;
CC GeneDB SPombe, SPAC26A3.16;
CC InterPro: IPR006436; ST1;
CC InterPro: IPR006449; UBA domain.
CC InterPro: IPR006426; Ubiquitin.
CC Pfam: PF00627; UBA; 1.
CC Pfam: PF00245; Ubiquitin; 1.
CC SMART: SM00727; ST1; 2.
CC SMART: SM00165; UBA; 1.
CC SMART: SM00213; UBO; 1.
CC PROSITE: PS00330; UBA; 1.
CC PROSITE: PS00053; UBIQUITIN_2; 1.
CC Hypothetical protein.
CC UBA DOMAIN 363 353 UBIQUITIN-LIKE.
CC UBA DOMAIN 363 353 UBA.
CC SEQUENCE 354 AA, 369.9 MW, 1A99E2D97E73A931 CRC64;

```

```

Query Match
Best Local Similarity 19.2%; Pred. No. 0.14;
Matches 80; Conservative 52; Mismatches 120; Indels 165; Gaps 15;

QY 39 SISLKLQIAGKLOESVDP--PELDLYCGRLKDDQTLDFYGIQNGTVVLRKSWPE 96
Db 25 SVLALKEAIA-----PVADIEKRCRLYAGRVLKDEESLRTYKTODGHSILHVKTLQGN 79
QY 97 P-----DQKP-----EPVDKYAAMREPRVLHLSHSSSY----- 126
Db 80 PAAATNVORTQCPVNIQAGGANNPLANLTSAR-YAGFNI MPMSASMGFPNPNPVP 138
QY 127 -----REAVFMLSNKSIEDQIIVATPGI-SSDPIALGVLOKDLFSVFA 170
Db 139 PSTELANLKNPNVQSSINEMFSNQMIDMINSPLHJRNAPPYVROMQSPERRANT 198
QY 171 DPNMLTTLVPAHVALVNAIVLHLSVAGSAPMPTDSSSRSPSSSSYRDMPGFLFEGLS 230
Db 199 DPQYFQMAQLHQQM-----GAAGIDPM-----SLMG-----GLG 229
QY 23: DDEDFHNTSTSSRPSASLYSQAAPRPIITOSLALALASPSSSHTPT 290

```

DE 230 G-----ACLGSGGAG-----240

QY 291 PTOGHSSGTSMSSGVSGTPTITLDFSLQALCHALQ-----A 328

DB 241 ---LGGFGANNATAGIAGAPVDCTAANTIONLXNLGGAGFGAGLGGAGLGGAG 297

QY 329 SCQPSLQSQWQP-----CLOQLPDMGLQDELSRRCRPVPTTSKQWSSSLLE 378

DB 298 ASPPAPACTRPPERYAEQLSQNEMGFVFFERNVQALRBS-IGNVQMA:ESLJSD 153

RESULT 6

EBN6_EBV

AC P03224; STANDARD; PRT: 992 AA

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last annotation update)

DE EBNA-6 nuclear protein (EBNA-3C) EBNA-4B.

GN BFR3-BERF4

OS Epstein-Barr virus (strain 955-8) (Human herpesvirus 4).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammaherpesvirinae; Lymphocryptovirus

OX NCBI_TaxID=10337;

RN (1)

RP SEQUENCE FROM N.A.

RA MEDLINE=84273667; PubMed=6287149;

RA Baer R., Bankier A.T., Biggin M.C., Deininger P.L., Farrell P.M., Gibson T.J., Hatfull G., Hudson G.S., Satchwell S., Segall C.E., Tuffnell P.S., Watling B.G.

RT "DNA sequence and expression of the 955-8 Epstein-Barr virus genome."

RL Nature 310:207-211 (1984).

RN (2)

RF CHARACTERIZATION.

RA MEDLINE=8835572; PubMed=2831334;

RA Petti L., Sample J., Wang F., Kieff E.

RT "A fifth Epstein-Barr virus nuclear protein (EBNA5) is expressed in latently infected growth-transformed lymphocytes."

RL J. Virol. 62:1330-1338 (1988).

RN (3)

RP SUBCELLULAR LOCATION.

RA MEDLINE=90266473; PubMed=2161150;

RA Petti L., Sample C., Kieff E.

RT "Subnuclear localization and phosphorylation of Epstein-Barr virus latent infection nuclear proteins."

RL Virology 176:561-574 (1990).

CC (1) FUNCTION: INVOLVED IN LATENT VIRUS

CC (2) SUBCELLULAR LOCATION: NUCLEAR

CC (3) SIMILARITY: SOME SIMILARITIES EXIST BETWEEN EBNA 4, EBNA 5, AND EBNA 6.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; V01555; CA24859.1;

DR Pfam; PF05009; EBNA-3; 1.

KW Nuclear protein; Repeat.

FT DOMAIN 74 80

FT POLY-APS

FT DOMAIN 551 610

FT 10 X 5 AA TANDYM REPEATS

FT DOMAIN 741 779

FT 3 X 11 AA TANDYM REPEATS

SQ SEQUENCE 992 AA; 109125 MW; 305EAB9B2515ED54 CWC47

Query Match 6.5%; Score 138; DB 1; Length 992

Best Local Similarity 22.5%; Pred No. 0.55;

Matches 71; Conservative 35; Mismatches 116; Indels 41; Gaps 15;

QY 177 TLVFAHPALVNAIVLHLS---VAGSAPMPGT---205

DB 461 TVEPAHTPVENPMWILLQPPVPVPVVKVTPFPSPRRRACVYDGLVETIDVETIE 520

DE

QY 206 DSSSRMPSSSYRDMFGGFLFGLSD-----DEDDFHPNTRSTPSSSTRPSALG 257

DB 521 DSSSVCPKPKRKHQDGFQFQSGRRKRAAPPTVSPDTPPAVGPPAAGPPAAGPAAG 580

QY 258 YSG-----AAGPRITSELTALALASTPESSSHPTPTGTQ-----294

DB 581 PPAAGPPAAGP-PAAGPRILAPLS-AGPPAAGPHIVTPPSARPRIVAPVVRMFMRQL 638

QY 295 GHSSGTPSVSS-GVQSGTPTINDLFSQALQHALQSCQPSL-QSQHQPQLQL--RDMG- 349

DB 639 POSTGRKPPQFWEXRAGREITQ--MQCEPSSHQSATQTPTPPSWAPSVCAISVMDACK 696

QY 350 ---IQDELSLRCPRPVGT*SKOPWSSSLLEEFH-----ELPASPEPPASCRGY-- 395

DB 697 AQPIESSHSSMSPTQIGHEQP-----RYEDCPAPLDLSLHPDVAAPQAPQVGYQE 752

QY 396 -----CPWEALMKVPP 406

DB 753 PPAPQAPVQYQVEPPP 768

RESULT 7

N153_RAT

AC P49791; STANDARD; PRT: 1468 AA

DT 01-OCT-1996 (Rel. 14, Created)

DT 01-OCT-1996 (Rel. 14, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Nuclear pore complex protein Nup153 (Nucleoporin Nup153) (153 kDa nucleoporin).

GN NUP153

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Rattus.

OX NCBI_TaxID=10116;

RN (1)

RP SEQUENCE FROM N.A. AND SEQUENCE OF 602-613; 622-645 AND 971-993.

RC STRAIN=Buffalo; TISSUE=Liver;

RX MEDLINE=93137325; PubMed=8422679;

RA Sukegawa J., Blobel G.

RT "A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and faces the nucleoplasm."

RL Cell 72:39-48 (1993).

CC (1) FUNCTION: POSSIBLE DNA-BINDING SUBUNIT OF THE NUCLEAR PORE COMPLEX (NPC). THE REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS OF THE PORE COMPLEX TO THE PORE MEMBRANE.

CC (2) SUBUNIT: INTERACTS WITH SERP2 (By similarity).

CC (3) SUBCELLULAR LOCATION: Nuclear pore complex; Located to the terminal rim structure of the nucleoplasmic cage.

CC (4) DOMAIN: Contains F-X-F-G repeats.

CC (5) PTM: Phosphorylated in interphase, hyperphosphorylated during mitosis. May play a role in the reversible disassembly of the nuclear pore complex during mitosis.

CC (6) SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST NUP1, NUP1, POX 121 AND MAMMALIAN P62.

CC (7) SIMILARITY: Contains 4 RANBP2-type zinc fingers.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; L06821; NOT ANNOTATED_CDS.

DR Pfam; A44345; A44345.

DR InterPro; IPR001476; Znf_RanGDP.

DR Pfam; PF03093; Nucleoporin PG; 9.

DR Pfam; PF00641; Znf_RanBP; 4.

DR SMART; SM00547; ZNF_PBZ; 4.

QY 355 LSLRPPCPVGTGKQPMSSSLDEHPHELPAPEPPASCRG 394
 DB 316 ---RP-EPPAGMASPPGGS-----PHELPKDQKSKALAG 346

RESULT 12

NR41 CANFA STANDARD: PRT: 598 AA.
 AC P51656;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2003 (Rel. 35, Last annotation update)
 DE Orphan nuclear receptor HMR (Orphan nuclear receptor; NRG1; NR4A1; OR HMR)
 OS Canis familiaris (Dog)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Carnivora; Fissipedina; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN 11;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97050607; PubMed=8895335;
 RA Pichon B., Jimenez-Cervantes C., Pirson I., Maenhaut C.,
 RA Christophe D.,
 RT "Induction of nerve growth factor-induced gene-B (NGFI-B) as an early
 RT event in the cyclic adenosine monophosphate response of doc
 RT thyrocytes in primary culture."
 RL Endocrinology 137:4691-4698(1996)
 CC -1- FUNCTION: Orphan nuclear receptor.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
 CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X97226; CAA65863.1;
 DR HSPSP; P19793; 2NLL;
 DR TRANSFAC; T04739;
 DR InterPro; IPR000536; Hormone_recep;
 DR InterPro; IPR001723; Steroid_recep;
 DR InterPro; IPR001628; Znf_C4steroid
 DR Pfam; PF00104; hormone_rec; 1
 DR Pfam; PF00105; zfc4; 1
 DR PRINTS; PR00398; SDRCHRMUNE;
 DR PRINTS; PR00047; STROCFINER;
 DR PRODOM; PD000035; Znf_C4ster; 1;
 DR SMART; SM00433; HCL1; 1
 DR SMART; SM00339; Znf_C4; 1
 DR PROSITE; PS00031; NUCLEAR_RECEPTO;
 KW Receptor; Transcription regulation; DNA binding; Nuclear protein;
 KW Zinc-finger; Phosphorylation.
 FT DNA_BIND 267 332 NUCLEAR RECEPTOR TYPE
 FT ZN_FING 267 287 C4-TYPE
 FT ZN_FING 303 327 C4-TYPE
 FT DOMAIN 409 459 LIGAND BINDING POTENTIAL
 FT DOMAIN 82 92 POLY-SER
 FT DOMAIN 583 586 POLY-PRO
 FT MOD_RES 341 341 PHOSPHORYLATION (BY PKA; BY SIMILARITY)
 FT MOD_RES 351 351 PHOSPHORYLATION (BY PKA; BY SIMILARITY)
 SQ SEQUENCE 598 AA; 64424 MW; 772563A8EC4F906 CRC64;

Query Match 6.0%; Score 129; DB 1; Length 598;
 Best Local Similarity 25.6%; Pred. No. 1;
 Matches 72; Conservative 26; Mismatches 119; Gaps 10;

QY 147 TPG-----LSSDPIALGVQLQ-CKDLFSVFAQRMCTILVPAHPALVNAIV --LVLHSA 197

DB 14 SPGRDHLASDPLTPELSKPTMDLASPEAAPT-APTALPSPSTFMDGTGTFDFTLYOLP 72
 QY 198 GSAPMPGTGSSSRSPSSS-----YRDMFG---GFLFEGLSDEDDDFH 237
 DB 73 GTAQPCSSASSASSASTSSSATSPASAFKFEEDFQVYCYCPGLSGPLDELTLSSGSDYY 132
 QY 238 PNTRSTPSSSTPSSRPASLG-YSGAAGRPRTITQSELATALALASTPSSSHPTTPGTQGH 296
 DB 133 GSPCSAPSPSTPSPFPQPLSPWDGSPFPSPFSQTYEGLRAWTEQLPKASGHPOPPAFFSF 192
 QY 297 SSGTSPYSSSWQSGTPT-INDLFSQALQHALCASQPSLQSQMOPQLQOLRDMG;ODDELS 356
 DB 193 SPPTGFSINLACSPULK-----FPSQATCOLGERESYSISTAPPG 232
 QY 357 LRCPRPVGTGSKQPMSSSLDEHPHELPAPEPPASCRGYCP 397
 DB 233 LAPTSFH-----LOGPGLMDA-PVPSAKARSGAP 260
 RESULT 13
 RA23.YEAST
 ID RA23 YEAST STANDARD: PRT: 398 AA.
 AC P32628;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE UV excision repair protein RAD23.
 GN RAD23 OR YEL037C OR SYGP-ORF29.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN 11;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94067136; PubMed=8246991;
 RA Watkins J.F., Sung P., Prakash S.,
 RT "The Saccharomyces cerevisiae DNA repair gene RAD23 encodes a nuclear
 RT protein containing a ubiquitin-like domain required for biolog:cal
 RT function."
 RL Mol. Cell. Biol. 13:7757-7765(1993).
 RN 12;
 RP SEQUENCE FROM N.A.
 RC STRAIN-B-6441;
 RX MEDLINE=94016559; PubMed=8411151;
 RA Melnick L., Sherman F.,
 RT "The gene clusters ARC and COR on chromosomes 5 and 10, respectively,
 RT of Saccharomyces cerevisiae share a common ancestry."
 RL J. Mol. Biol. 233:382-388(1993).
 RN 13;
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288c / AB 97.
 RX PubMed=9169868;
 RA Dietrich F.S., Mullen J.T., Hennessy K.M., Yelton M.A., Allen E.,
 RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
 RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartrell G.,
 RA Huntic-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
 RA Lin D., Moseedale D., Nakahara K., Namath A., Norgren R., Oetner P.,
 RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
 RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.,
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V."
 RL Nature 387:78-81(1997)
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: Contains 1 ubiquitin-like domain.
 CC -1- SIMILARITY: Contains 2 UBA domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces;
NCBI_TaxID=4896;

[...]
SEQUENCE FROM N.A.

STRAIN=972;

MEDLINE=21848401; Pubmed=11819360;

Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,
RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

RA Wood V., Gwallier R., Rajadurai M.A., Lyne M., Lyne R., Stewart A.,

FT DOMAIN 375 378 POLY-SER.
SQ SEQUENCE 565 AA; 59713 MW; F56CCCF7E54A8CDB CRC64;

Query Match 5.9%; Score 126; DB 1; Length 565;
Best Local Similarity 22.38; Pred. No. 1.4;
Matches 82; Conservative 47; Mismatches 138; Indels 100; Gaps 16;

QY 74 QTLDFYGI:-----QPGSTVHVJRKSWPEPDQKPEVQKVAAMREFRVLHTALHSSSY 126
DB 164 OPATIVGIPKALHTTONGTTTHVGTTPGAVSQSRSEPDRLSSNMVQMLASA---AAQ 240

QY 127 REAVFKYLSKESLDQIVATPGHSSDPALGVLODKLFSVADPNMIDTLVPAHALY 186
DB 241 LCAAPRICITKSS:-----GVNLMPLSNAPSPKQMNYYGSLPSSNT-SPNHLASV 290

QY 187 NAIIVLVHSAVAGSAPMPGDTSSSRMPSSSVYRWMPGGFLFEGLSDDCEDFHP-----NTR 241
DB 291 PNRGLTNSSTGSEF-TKYNSSNSLYNSNMOTP---YLPKSNSTSLHSMYGVTTA 346

QY 242 STPSS:-----STPSSRPA-SLOYSGAAG-----PRPITQSELTALALAST 281
DB 347 YAPQSLRYAHNYLPSRPSVNGFDODSSSDFAHFRYQRRSRPVSPCTAPSSPTFT 406

QY 282 PESSHTTETG:-----TQHSSTGTPMSSGVQSGTPTINDLFSQALQALQASQPSLOS 336
DB 407 ---RSFSPITDVTPLVTPAHSPRURPVD-----DPSC-- 435

QY 337 QWQPLQCLPDMGICQDELSLRP---CRPPVGTGKQPMSSSLLEPHELPA: 385
DB 436 ---VQLPILSLSRPSQVP::PPLKCDPNAFSASTPASGAVSRTPSSVLSLSLVNVS 492

QY 386 PEPPAS 39:
DB 493 MPKPEAS 499

Search completed: November 12, 2003, 06:37:43
Job time : 20 secs

Craniata; Vertebrata; Euteleostomi;
Catarrhini; Homnidae; Homo.

SQ SEQUENCE 380 AA; 40510 MW; 08649F79597E0931 CRC64;
 Query Match 86.0%; Score 1834; DB 4; Length 380;
 Best Local Similarity 130.0%; Pred. No. 35e-129;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLSQWHLAVKLAQOPLTPKSIILRPETELGEYSLSGYSISFLKQLIAGKLGQESVDPDEL 60
 DB 1 MSLSQWHLAVKLAQOPLTPKSIILRPETELGEYSLSGYSISFLKQLIAGKLGQESVDPDEL 60
 QY 61 IDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSWPEPQKPEPVCKVAAMREFVLHTAL 120
 DB 61 IDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSWPEPQKPEPVCKVAAMREFVLHTAL 120
 QY 121 HSSSYREAVFKMLSNKESLDQIIIVATPGLSSDPALGVLOKDLFSVFADPNMLDTLVP 180
 DB 121 HSSSYREAVFKMLSNKESLDQIIIVATPGLSSDPALGVLOKDLFSVFADPNMLDTLVP 180
 QY 181 AHPALVNAIVLVHVSAGSAPMPGTDSRSSRMPSSSYRDMPGGFLFEGLSDDDDPHNT 240
 DB 181 AHPALVNAIVLVHVSAGSAPMPGTDSRSSRMPSSSYRDMPGGFLFEGLSDDDDPHNT 240
 QY 241 RSTPSSSTPSSRPASLGYSAGAPRPITQSELATALALASTPSSSHHTPTPTQGHSSGT 300
 DB 241 RSTPSSSTPSSRPASLGYSAGAPRPITQSELATALALASTPSSSHHTPTPTQGHSSGT 300
 QY 301 SPMSGVSQSGTPTINDLFSQALQHALQASGQPSLOSQWQPQLQQLRDMGIDDEL 358
 DB 301 SPMSGVSQSGTPTINDLFSQALQHALQASGQPSLOSQWQPQLQQLRDMGIDDEL 358

RESULT 2
 Q96S82 PRELIMINARY; PRT; 380 AA
 AC Q96S82;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Ubiquitin-like protein S5.32.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catartini; Hominoidea; Hominidae; Homininae;
 OC NCBI_TaxID=9606;
 RN 1;
 RP SEQUENCE FROM N.A.
 RA Zhang W., Li N., Wan T., Cao X.;
 RA Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AY037166; ANK67643.1;
 DR InterPro; IPR0019C1; SecE;
 DR InterPro; IPR000649; UBA domain;
 DR Pfam; PF00627; UBA; 1;
 DR SMART; SM00165; UBA; 1;
 DR SMART; SK00213; UBA; 1;
 DR PROSITE; PSC1067; SEC6; SEC6G; 1;
 DR PROSITE; PS00053; UBIQUITIN 2; 1;
 SQ SEQUENCE 380 AA; 40581 MW; 5AF6A14F81F996C CRC64;

Query Match 85.5%; Score 1824; DB 4; Length 380;
 Best Local Similarity 99.4%; Pred. No. 2e-128;
 Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSLSQWHLAVKLAQOPLTPKSIILRPETELGEYSLSGYSISFLKQLIAGKLGQESVDPDEL 60
 DB 1 MSLSQWHLAVKLAQOPLTPKSIILRPETELGEYSLSGYSISFLKQLIAGKLGQESVDPDEL 60
 QY 61 IDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSWPEPQKPEPVCKVAAMREFVLHTAL 120
 DB 61 IDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSWPEPQKPEPVCKVAAMREFVLHTAL 120
 QY 121 HSSSYREAVFKMLSNKESLDQIIIVATPGLSSDPALGVLOKDLFSVFADPNMLDTLVP 180

DB 121 HSSSYREAVFKMLSNKESLDQIIIVATPGLSSDPALGVLOKDLFSVFADPNMLDTLVP 180
 QY 181 AHPALVNAIVLVHVSAGSAPMPGTDSRSSRMPSSSYRDMPGGFLFEGLSDDDDPHNT 240
 DB 181 AHPALVNAIVLVHVSAGSAPMPGTDSRSSRMPSSSYRDMPGGFLFEGLSDDDDPHNT 240
 QY 241 RSTPSSSTPSSRPASLGYSAGAPRPITQSELATALALASTPSSSHHTPTPTQGHSSGT 300
 DB 241 RSTPSSSTPSSRPASLGYSAGAPRPITQSELATALALASTPSSSHHTPTPTQGHSSGT 300
 QY 301 SPMSGVSQSGTPTINDLFSQALQHALQASGQPSLOSQWQPQLQQLRDMGIDDEL 358
 DB 301 SPMSGVSQSGTPTINDLFSQALQHALQASGQPSLOSQWQPQLQQLRDMGIDDEL 358

RESULT 3
 Q91W67 PRELIMINARY; PRT; 380 AA
 AC Q91W67;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Similar to RIKEN cDNA 2300004C15 gene.
 GN 2300004C15RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN 1;
 RP SEQUENCE FROM N.A.
 RA Tissue=Salivary gland;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL: BC016456; AAH16456.1;
 DR MGD; MG:1916709; 2300004C15RIK.
 DR InterPro; IPR0019C1; SecE;
 DR InterPro; IPR000649; UBA domain;
 DR InterPro; IPR000626; Ubiquitin;
 DR Pfam; PF00627; UBA; 1;
 DR Pfam; PF00240; Ubiquitin; 1;
 DR SMART; SM00165; UBA; 1;
 DR SMART; SM00213; UBA; 1;
 DR PROSITE; PSC1067; SEC6; SEC6G; 1;
 DR PROSITE; PS00053; UBIQUITIN 2; 1;
 SQ SEQUENCE 380 AA; 40407 MW; 6DE0834A7CCEARF2 CRC64;

Query Match 84.6%; Score 1804; DB 11; Length 380;
 Best Local Similarity 97.9%; Pred. No. 6.1e-127;
 Matches 350; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSLSQWHLAVKLAQOPLTPKSIILRPETELGEYSLSGYSISFLKQLIAGKLGQESVDPDEL 60
 DB 1 MSLSQWHLAVKLAQOPLTPKSIILRPETELGEYSLSGYSISFLKQLIAGKLGQESVDPDEL 60
 QY 61 IDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSWPEPQKPEPVCKVAAMREFVLHTAL 120
 DB 61 IDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSWPEPQKPEPVCKVAAMREFVLHTAL 120
 QY 121 HSSSYREAVFKMLSNKESLDQIIIVATPGLSSDPALGVLOKDLFSVFADPNMLDTLVP 180
 DB 121 HSSSYREAVFKMLSNKESLDQIIIVATPGLSSDPALGVLOKDLFSVFADPNMLDTLVP 180
 QY 181 AHPALVNAIVLVHVSAGSAPMPGTDSRSSRMPSSSYRDMPGGFLFEGLSDDDDPHNT 240
 DB 181 AHPALVNAIVLVHVSAGSAPMPGTDSRSSRMPSSSYRDMPGGFLFEGLSDDDDPHNT 240
 QY 241 RSTPSSSTPSSRPASLGYSAGAPRPITQSELATALALASTPSSSHHTPTPTQGHSSGT 300
 DB 241 RSTPSSSTPSSRPASLGYSAGAPRPITQSELATALALASTPSSSHHTPTPTQGHSSGT 300
 QY 301 SPMSGVSQSGTPTINDLFSQALQHALQASGQPSLOSQWQPQLQQLRDMGIDDEL 358
 DB 301 SPMSGVSQSGTPTINDLFSQALQHALQASGQPSLOSQWQPQLQQLRDMGIDDEL 358

RESULT 4

Q9D7P5 PRELIMINARY; PRT; 380 AA.
 AC Q9D7P5;
 DT 01-JUN-2001 (TRENBLREL. 17, Created)
 DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE 2300004C15Rik protein.
 GN 2300004C15Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RC MEDLINE=21085660; PubMed=1127851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Kanno H., Adachi C., Fukuda S.,
 Aizawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kaeska K., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kouchiwa M.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido T., Pevsner J., Quackenbush J.,
 Schriml L.M., Staab J.F., Suzuki R., Tomita M., Wagner D., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo V.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi V.,
 Gusticich S., Hill D., Hofmann V., Hume D.A., Kariya W., Lee K.H.,
 Lyons P., Marchionni L., Mashima M., Mazzarelli L., Mombaerts P.,
 Nordone P., King B., Ringwald M., Rodriguez L., Sakamoto N.,
 Saeki H., Sato K., Schoenbach K., Seya T., Shibata Y., Storch K.F.,
 Suzuki H., Toyooka K., Wang K.H., Weitz M., Whitaker C., Wilting L.,
 Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschütter S.,
 Hayashizaki Y.
 RA "Functional annotation of a full-length mouse cDNA collection."
 RT Nature 409:685-690(2001).
 RL Nature 409:685-690(2001).
 DR EMBL; AK090025; BAB26033.1;
 DR MGD; MG11916709; 2300004C15R.k.
 DR InterPro; IPR001901; SecE.
 DR InterPro; IPR000449; UBA domain.
 DR InterPro; IPR000626; Ubiquitin.
 DR Pfam; PF00627; UBA; 1.
 DR Pfam; PF00240; Ubiquitin; 1.
 DR SMART; SM00165; UBA; 1.
 DR SMART; SM00237; URQ; 1.
 DR PROSITE; PS01067; SECE; SEC6G; 1.
 DR PROSITE; PS50953; UBQUITIN_2; 1.
 DR PROSITE; PS50953; UBQUITIN_2; 1.
 SC SEQUENCE 380 AA; 40410 MW; 2300004C15Rik.
 Query Match 93.5%; Score 188.5; DB 11; Length 380.
 Best Local Similarity 96.6%; Pred. No. 8.2e-06.
 Matches 146; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
 QY 1 MSLSDHVLAVKADQPLTPKSLIPETELGYSLSFLKQLIAGKLOESVPPDEL 60
 DB 1 MSLSDHVLAVKADQPLTPKSLIPETELGYSLSFLKQLIAGKLOESVPPDEL 60
 QY 61 IDLIYCGKRLKDDQTLDFYGIQPGSTVHVLKSKMPETCKPEPVKVAAMKEFFVLHTAL 120
 DB 61 IDLIYCGKRLKDDQTLDFYGIQPGSTVHVLKSKMPETCKPEPVKVAAMKEFFVLHTAL 120
 QY 121 HSSSYREAVFPMNLNKSLSQIIIVATPGLSSDIIAGLQCKLFSVFADPNVLTLP 180
 DB 121 HSSSYREAVFPMNLNKSLSQIIIVATPGLSSDIIAGLQCKLFSVFADPNVLTLP 180
 QY 181 AHPALNVAIVLVJHSVAGSAPYPTGDSRSSYSSYRDMEGGLFEGLSDDDECHPNT 240
 DB 181 AHPALNVAIVLVJHSVAGSAPYPTGDSRSSYSSYRDMEGGLFEGLSDDDECHPNT 240
 QY 241 RSTPSSSTPSSRPASGLYSGAAGRPITQSELATALALASTPSSSHTPTGTGQHSSTG 300

DB 241 RSTPSSSTPSSRPASGLYSGAAGRPITQSELATALALASTPSSSHTPTGTGQHSSTG 300
 QY 301 SPMSGVSQSGTPIITNDLFSQALQASGOPSLSQSQWQLOQLQRLDMGTQDDLSLR 358
 DB 301 SPMSGVSQSGTPIITNDLFSQALQASGOPSLSQSQWQLOQLQRLDMGTQDDLSLR 358
 RESULT 5
 Q8K141 PRELIMINARY; PRT; 554 AA.
 ID Q8K141;
 DT 01-OCT-2002 (TRENBLREL. 22, Created)
 DT 01-OCT-2002 (TRENBLREL. 22, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RA Strausberg R.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; SC028857; AAH28857.1;
 DR InterPro; IPR000449; UBA domain.
 DR InterPro; IPR000626; Ubiquitin.
 DR Pfam; PF00627; UBA; 1.
 DR SMART; SM03165; UBA; 1.
 DR SMART; SM03213; URQ; 1.
 DR PROSITE; PS50053; UBQUITIN_2; 1.
 DR Hypothetical protein.
 SC SEQUENCE 554 AA; 58677 MW; P255C9985A4C029 CRC64;
 Query Match 8.8%; Score 188.5; DB 11; Length 554.
 Best Local Similarity 23.0%; Pred. No. 8.2e-06.
 Matches 104; Conservative 76; Mismatches 174; Indels 99; Gaps 21;
 QY 15 QPLTPKSLIPETELGYSLSFLKQLIAGKLOESVPPDELIDLYCGKRLKDD 73
 DB 24 EPKIMKVTGTPK-EKEEFVAVPSSVQFKEEISKRFKSHI---DQLVLI-FAGKILKDD 79
 QY 74 QTLDFYGIQPGSTVHVLKSKMPETCKPEPVKVAAMKEFFVLHTAL 112
 DB 80 DTLSDHVLAVKADQPLTPKSLIPETELGYSLSFLKQLIAGKLOESVPPDELIDLYCGKRLKDD 112
 QY 113 FRV-----LHTALHSSSYREAVFPMNLNKSLSQIIIVATPGLSSDIIAGLQCKLFSVFADPNVLTLP 160
 DB 138 FGVGLGGIAGLSLSGLNTTNFSELSQSQWQLOQLQRLDMGTQDDLSLR 190
 QY 161 QKDLFS-VFADPNVLTLPVHVLKSKMPETCKPEPVKVAAMKEFFVLHTAL 213
 DB 191 SNPDLMLQINANPOM-QOLIQRNPEISHMLNPNDRQTLLEARNPAMMOEMNRNDRA 249
 QY 214 SSSYRDMEGGLFEGLSDD-----EDFHPNTRSTPSSSTPSSRPASGLYSGAAG 263
 DB 250 LSNLESIFGYNALRMVTDIQEPLNAAQOFGN----PFASLVSS-----SBAEG 299
 QY 264 PRPITQSELATALALASTPSSSHTPTGTGQHSSTGTPMSGVSQSGTPIITNDLFSQALQ 323
 DB 300 TQP-SRTENPDLNPNWAPQSQSPASGTTGTTNTMTSGGTATSTPAQSQSGPSLV 358
 QY 324 HALQAS--GQPSLSQSQWQ-----POLQQLRDMGTQDDLSLRPCRPV-GTSKQPSVSL 375
 DB 359 PGAGASMTNPGMSLLOOITENPOLMQ-----NMLSAPYRSMQLQSLQNPDLAAQ 410
 QY 376 LEEPELFPASPEPPACRCGCPWEALMKVPPSL 408
 DB 411 MNPDTLSMNSNRA-----MQALLIQOGL 436
 RESULT 6


```

RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carnacci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Kallin-Neumann G., Kawai J., Kim C., Lam B., Lin C.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shirai P., Southwick A., Shinzaki K.,
RA Davis R.W., Ecker J.R., Theologis A.,
RT "Arabidopsis Open Reading Frame (ORF) Clones."
RL Submitted (MAY-2002) to the EMBL/GenBank/CDR databases.
DR EMBL: AC007127; AA225137.2;
DR EMBL: AF362159; AA25869.1;
DR EMBL: AY042828; AA68768.1;
DR EMBL: AY081450; AA10012.1;
DR EMBL: AY113886; AA44934.1;
DR InterPro: IPR000443; UBA domain.
DR Pfam: PF00627; UBA;
DR Pfam: PF02240; ubiquitin;
DR PROSITE: PS00053; UBIQUITIN-2;
SQ SEQUENCE 538 AA: 57285 MW: 47286 P25F36413RBD CR-647

Query March 7.78; Score 164.5; DR 10; Length 538;
Best Local Similarity 24.08; Pred. No. 0.0042;
Matches 99; Conservative 49; Mismatches 146; Indels 125; Gaps 19;

QY 39 SISFLKQIAGKQESVDPDELIDLYGKQKQKQDQDTDFYGTQGTSTVYVLPKSWPEPD 98
DQ :|||:
DB 39 TVESFKELIA---CNSDVPANQCRUYKGRILKDDCTLLSYGCAQHTVMYVPGFVSSP 95
QY 99 QKP-----EPVKVAYREP--- --AV 115
DB 96 SAPAANAGNOTTAPQAVGSSNDSNGESLFDGPNPLGGNAGLFGSLDPLECA 155
QY 116 LHTALHSSSYRE---AVFKLSNRES-DQIVATRLSSNPLALGVLOCK- DLRV 149
DB 156 QQGLAQNPNWIREMNTPTAIONLXNPEYRSNKNPQVRE -----LVGNELGHV 208

169 FADPNM-DTLVPA-HPALVNA-VLVLSHVSAGAPMPTDSSSPNMPSSSYDNGPF- 224
209 LNDSEILQGLEARNPELRM- -----MENTD- PAM-SNIRSYEPGNM 251
QY 225 ---LFEGLSDDEDFHNTSTSTSTPSRF---ASQVSGAGAPRPITCSKATALALA 279
DB 252 LRMYENVY-CEPLXNATTXSNAGNTSNRIALLNGVTTQSDTSNNAIPNAET 309
QY 280 STPESSHTPTQGHSSG--TSVSSGVCSG- --- --TFITN 315
DB 310 GTPVAN --FLPNWQATA-CTATNENAGLILALILALILALILALILALILALILAL 365
QY 316 DLFSQALCHLQASVQSLVAVG- --- --LQVAVVAVVAVVAVVAVVAVVAVVAVVAV 367
DB 367 QL-SQLLNTFAMSGVSVASNRYWMLQALNATLQVLEWMLQVLEWMLQVLEWMLQV 418

RESULT 13
QY Q9VWD9 PRELIMINARY; PRT, 547 AA.
AC Q9VWD9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created);
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update);
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update);
DE CG14224 protein (CG38919p).
GN CG14224.
OS Drosophila melanogaster (Fruit fly);
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;

```

```

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazei R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.D.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Bencs P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup E.E., Downes M., Durkin-Koch S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Fertiz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekurov G., Mishra N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb C.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinett K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson V., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector K., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang X.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye C., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RC STRAIN=Berkeley;
RA Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gorzalez R., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungal C.J., Nunoo C., Pacleb C., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.,
RL Submitted (JUL-2002) to the EMBL/GenBank/CDR databases.
DR EMBL: AC003512; AA549003.1;
DR EMBL: AY128447; AA79540.1;
DR HSSP: P02248; 13E1;
DR Flybase: FBgn0031057; CG14224;
DR InterPro: IPR000443; UBA domain.
DR InterPro: IPR00626; ubiquitin.
DR Pfam: PF00627; UBA;
DR Pfam: PF02240; ubiquitin;
DR PROSITE: PS00053; UBIQUITIN-2;
SQ SEQUENCE 547 AA: 58836 MW: 84960 C6EC165C4D7 CRC64;

Query March 7.68; Score 161.5; DB 5; Length 547;
Best Local Similarity 22.48; Pred. No. 0.00085;
Matches 100; Conservative 60; Mismatches 170; Indels 117; Gaps 20;

QY 18 TPKSILKLPETELGEYSLOGYSISFLKQIAGKQESVDPDELIDLYGKQKQDQDTLD 77
DQ :|||:
DB 15 TPKD---KKTVEVDEDS---GIMDKILVAQKE---AEPEQLVLIAGKIMKDTDLQ 64
QY 78 FYGIQPGSTGVHLKSKSPEDQKE--EPVD-----KVAAMR----- 111
DB 65 MHNKIDNLTVHLVTKATRNNECPARQPADVROTPTGLNKGFLGALGAGSNTFMDL 124
QY 112 EFRVLHTALHSSSYRE-----AVFKLSNRESLDQIVATPG----LSSDPTALGVLD 162
DQ :|||:
RX MEDLINE=20196006; PubMed=10731132;

```

```

DE 125 QARMONELLINGDMLSRNDYDPMVQMMNNEFTMRQLTSRPMQHDLMQNPESHVLYN 184
QY 163 KDL-----FSVFADPNMLDTCPAHP-ALVNA:VLVLSVAG -----SAPM--P 203
DB 185 PDLLEQTVELARNPMLQELNRSHPAMN- LSVPGYSALQRIYRDQERPMKA 239
QY 204 GTDSSRSMPASSYRDMPPGGFLFEGLSDDECFHFNTRSTPSSSPSSPASLYSGAAG 263
DB 240 ATESFGRN PFAGLV--GGSGAGNNPQCCOTENRNPLFNFWNANSNTNVTGSGGAKN 296
QY 264 P-----EPITQSELATALALATPSSSHITPTGTOHSGSTSPMSGVSQGTPIFNLFQALQ 323
DE 297 PTGELPPNVLNTPAIVSLQVNAIPAMQRLNARYTRSMYSMSQQLMAAPLS - 354
QY 300 TSPMSGVSQGTPIFNLFQALQVNAIPAMQRLNARYTRSMYSMSQQLMAAPLS 358
DB 355 -----SEPLMSNPAIQEQVQXVTFNAGMGNFEMVWVLTNFIAMNALLQIQIMVQPS 410
QY 348 M--GIQDELSLRRCPRPVGTSKQWS 312
DB 411 AAPGLV GTLGIPPPGAGTGNDPAS 416

RESULT 14
Q9JUP9 PRELIMINARY; PRT: 582 AA.
AC Q9JUP9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created;
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE DA41.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 1;
RP SEQUENCE FROM N.A.
RC Tissue-Lung;
RA Ozaki T., Hishiki T., Nakagawara A., Sakiyama S.
RT "Identification of a new protein that can associate with LAM1."
RL Submitted (SEP 1996) to the EMBL/GenBank/CCDS databases.
DR EMBL: D87950; BAA92267.1;
DR HSSP: P02248; ITB6;
DR InterPro: IPR006636; STI1;
DR InterPro: IPR000449; UBA domain.
DE InterPro: IPR000626; Ubiquitin.
DF Pfam: PF00627; UBA; 1.
DR Pfam: PF00240; ubiquitin; 1.
DR SMART: SM00727; STI1; 4.
DR SMART: SM00165; UBA; 1.
DR SMART: SM00213; UEQ; 1.
DR PROSITE: PSS0053; UBIQUITIN 2; 1.
SQ SEQUENCE 582 AA; 62071 MW; 455,858 Da; 1000 CP; 1000

Query Match 7.4%; Score 158.5; DB 11; Length 582;
Best Local Similarity 22.7%; Pred. No. 0.0015;
Matches 88; Conservative 67; Mismatches 150; Indels 83; Gaps 18;

QY 15 QPTPKSLRLPETELGEYSLGGYSYSLFKQIAGKQESVDPPELIDIVYQKLD 73
DB 24 EPKIMKVTYTPK-EKEEFAPVENSVCQFKELSKFKSQ--DQLVLFAGKILQ 79
QY 74 QTLDFGIQGTSTVHLRK-----SAPPEQCFEPVQVNAEE 112
DB 80 DTLRSQGHDEGLTVHLVITKQRPQENSACQNTTNSVTSPPAES--NTSGPAANS 137
QY 113 FRV-----LHTALHSSSSVREAVFM--LKKESLDLIIVATPGLSSDPALGV 160
DB 138 FLGGGLAGLSSGLNTTNSFELCSNQRCQRLNENPMKLN-----ENFVQSKL 190
QY 161 QDKDLFS--VFADPNMLDTLVAHPALVNA--IVLVLSVAGSAMPPTGSSSRXP 213
DB 192 SNPDLMQRLMANPQV-QQLIQNPFIHSHVLYNPN:MQTLEARNPANYQEMVQNRQD 249

```

```

QY 214 SSSYRDMPPGGF-----LFEGLSD-----EDDFHFNTRSTPSSSPSSPASLYSGAAG 263
DB 250 LSNLESIQGYNALRMYTDIQEPLNAAQEQFGN-----PFASLVSP-----SSNRG 299
QY 264 PRPTQSELATALALATPSSSHITPTGTOHSGSTSPMSGVSQGTPIFNLFQALQ 323
DB 300 TQP-SRTENRDPLEPFWAPQTPQSSPASGSGTGTNTTSTAGNATSTPAQGTSGPNLV 358
QY 324 HALQAS--GQPSLOSQWQ-----PGLQQ 344
DB 359 PGAGASMPNTPGMSLQOQITENPOLYC 386

RESULT 15
Q9UHD9 PRELIMINARY; PRT: 624 AA.
AC Q9UHD9;
DT 01-MAY-2000 (TrEMBLrel. 14, Created;
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Ubiquitin-like product Chap./Dsk2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=20141258; PubMed=10675567;
RA Kaye F.J., Modi S., Ivanovska L., Koonin E.V., Thress K., Kubo A.,
RA Kornbluth S., Rose M.D.;
RT "A family of ubiquitin-like proteins binds the Affase domain of Hsp70-
RT like Stch."
RL FEBS Lett. 467:345-355(2000).
DR EMBL: AF189509; AAF17237.1;
DR HSSP: P02248; ITB6;
DR InterPro: IPR006636; STI1;
DR InterPro: IPR000449; UBA domain.
DE InterPro: IPR000626; Ubiquitin.
DF Pfam: PF00627; UBA; 1.
DR Pfam: PF00240; ubiquitin; 1.
DR SMART: SM00727; STI1; 4.
DR SMART: SM00165; UBA; 1.
DR SMART: SM00213; UEQ; 1.
DR PROSITE: PSS0053; UBIQUITIN 2; 1.
SQ SEQUENCE 624 AA; 65765 MW; 45753850 Da; 66AC3 CRC64;

Query Match 7.1%; Score 152.5; DB 4; Length 624;
Best Local Similarity 19.8%; Pred. No. 0.0048;
Matches 97; Conservative 70; Mismatches 177; Indels 147; Gaps 19;

QY 13 AQPILTPN-----LRLPETELGEYSLGGYSYSLFKQIAGKQESVDPPELIDIVYQGR 68
DB 24 AAAPAEPKIKVTYTPK-EKEEFAPVENSVCQFKELSKFKSQ--DQLVLFAGK 79
QY 69 KLKDDQTLDFYGFQGTSTVHLRKSHPEPD-QKPEVDVNAVREFRVLHTALHSSSVR 127
DB 80 ILKDDQTLDFYGFQGTSTVHLRKSHPEPD-QKPEVDVNAVREFRVLHTALHSSSVR 130
QY 128 EAVFMKLSN-----KESLDLIIVATP-----GLSSDPI 155
DB 131 SNSTP-STNSNPFGLSLQGLAGLSLGLSTNFSLOSOQOOLMASPEMIIQIMENPF 190
QY 156 ALGVQCKDLFS--VFADPNMLDTLVAHPALVNA-----IVLVLSVAGSAMPPTGSS 208
DB 191 VQSMLSNPOLMQLINANPQV-QQLIQNPFIHSHVLYNPN:MQTLEARNPANYQEMVQNR 249
QY 209 SRSMPSYRDMPPGGF-LFEGLSDDEDFHFNTRSTPSSSPSSPASLYSGAAG----- 263
DB 250 NQDLALSNLESTPGGYNALRRMTD:QEPMLNAAQEQFGNPF-----ASVSSSSSGGTQ 306
QY 264 -----PRPTQSELATALALATPSSSHIT-----P 289

```


Db	307	PSRTNRDPLNPMAPPATQSS-ATSTTTTSGSGNSSSNATONTVAANYVASIFS	365
Qy	290	TPGTQG-----HSSGTSPMSSGVQSGPITNDLFSQALQHA-QASGQPSLQSQ	337
Db	366	TPGVCSILQQITENPQLIQKXLSAPYKSMWQSLSQNFDLAQQYMLNSPLFTANPQLQEQ	425
Qy	338	WQPOLQQLRDMGQDDELSURPQPPVGTSGKQFWSSSLEEEHEUPASPEPPASGQY.P	397
Db	426	MRPQL-----EAFQQMONTTUSAYNPPA	451
Qy	398	WEALYKYVPPSL	408
Db	452	MQALYQIQQGL	462

Search completed: November 12, 2003, 04:04:34
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: Nov-Dec 12, 2003, 06:12:23 : Search time 46 seconds
without alignment
14:41:37 Million cell updates/sec

Title: US-09-735-251-4

Perfect score: 2133

Sequence: 1 MSLSHMLAVKLADQLTPK ... SGPVYCVKPMKVFYV 410

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 159724573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A Geneseq 19Jun03

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2004.DAT
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2005.DAT
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2006.DAT
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2007.DAT
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2008.DAT
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2009.DAT
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2010.DAT
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2011.DAT

Pred. No. is the number of results predicted by search to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2133	100.0	410	18 AAW15414	Activated platelet
2	2129	99.8	426	18 AAW15413	Human activated pi
3	1834	86.0	420	20 AAY42182	Amino acid sequenc
4	1834	86.0	420	22 AAU39074	Human secreted pro
5	1834	86.0	420	23 ABB55523	Human polypeptide
6	1828	85.7	380	22 AAG31352	Human protein HPLC
7	1686	79.0	341	23 ABB84954	Human polypeptide
8	794	37.2	191	22 AAB61348	Human prostaglan
9	592	27.8	164	22 AAB61349	Human prostaglan

10	566	26.5	128	22 AAO13489	Human polypeptide
11	389	18.2	76	21 AAG02759	Human secreted pro
12	370	17.3	114	21 AAB58780	Breast and ovarian
13	164.5	7.7	536	21 AAG30040	Arabidopsis thalia
14	164.5	7.7	577	21 AAG30039	Arabidopsis thalia
15	161.5	7.6	547	22 ABB59125	Drosophila melanog
16	152.5	7.1	624	22 AAB47122	Human Chapi. Homo
17	152.5	7.1	624	23 ABB97173	Novel human protei
18	152	7.1	329	23 ABB10042	Yeast selected int
19	149	7.0	415	22 ABB59727	Drosophila melanog
20	145.5	6.8	589	22 AAB94008	Human stomach canc
21	145.5	6.8	589	22 AAB95035	Human protein sequ
22	141.5	6.6	589	22 AAB94311	Human protein sequ
23	141.5	6.6	589	22 AAB62194	Human SAP2 amino a
24	141.5	6.6	604	22 ABB16557	Novel human diagno
25	140.5	6.6	589	22 ABB32717	Human brain-derive
26	139.5	6.5	1093	16 AAB66460	AF-17 protein. Ho
27	139.5	6.5	1093	23 ABB95048	Human translocatio
28	139	6.5	589	22 AAB74670	Human protease and
29	136.5	6.4	464	21 AAB12447	Human 149Y2H151 u
30	134	6.3	655	21 AAY85624	Human Zalpha16 ami
31	134	6.3	655	24 ABB73191	Human Zalpha16 pro
32	134	6.3	655	24 ABB73251	Human cancer/testi
33	134	6.3	1145	22 AAB48895	Microsporidia eve
34	134	6.3	1963	22 ABB62919	Drosophila melanog
35	133.5	6.3	902	24 ABB54281	Human expressed pr
36	133.5	6.3	902	24 ABB54281	Human expressed pr
37	133.5	6.3	907	24 ABB94278	Human expressed pr
38	133.5	6.3	907	24 ABB94278	Human expressed pr
39	133.5	6.3	907	24 ABB94278	Human expressed pr
40	133.5	6.3	2311	23 ABB95034	Human translocatio
41	133	6.2	1012	20 AAY17406	Human atrophin-1 r
42	133	6.2	1168	22 ABB02040	Novel human diagno
43	133	6.2	1487	23 AAO14356	Equine herpesvirus
44	132	6.2	4899	22 ABB65895	Drosophila melanog
45	131.5	6.2	773	22 ABB15668	Novel human diagno

ALIGNMENTS

RESULT 1

AAW15414
ID AAW15414 standard; Protein: 410 AA.

XX AAW15414

AC AAW15414

DT 06-JUN-1997 (first entry)

XX Activated platelet protein-2 APP-2 alternatively spliced variant.

XX APP-2; activated platelet protein-2; thrombus; monoclonal antibody.

XX Homo sapiens.

XX WO9712898-A1.

XX 10-APR-1997.

XX 04-OCT-1996; 96WO-US15922.

XX 06-OCT-1995; 95US-0005074.

XX (HARD) HARVARD COLLEGE.

XX Clement CY, Reed G;

XX WPI; 1997-226153/20.

XX N-PSDB; AAT64550.

XX Naturally-occurring platelet activation polypeptide, APP-2 - used

PT for locating a platelet thrombus in an animal, or activated platelet

PT complex in a biological sample

XX Claim 2; Fig 6; 50pp; English.

PS of novel human activated platelet protein-2 (APP-2) (see also

XX AA15413), a protein that is preferentially expressed on activated

CC human platelets but not resting platelets. Its amino acid sequence

CC was deduced from an isolated cDNA clone (AA15450). APP-2 is

CC characterised as confg. an epitope that binds to monoclonal antibody

CC 3B2 (ATCC CRL 11986). APP-2 can be produced in transformed host

CC cells. Anti-APP-2 antibodies are useful in methods for detecting

CC activated platelets in biological samples, for localising a platelet

CC thrombus in an animal, and for detecting an activated platelet

CC protein complex in an animal.

XX SQ Sequence 410 AA;

Query Match 100.0%; Score 2132; DB 18; Length 410;

Best Local Similarity 100.0%; Pred. No. 5.1e-167;

Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLSDWHLAVKLAADQPLTPKSIILRLPETELGYSLSGGYSISFLKQIAGKQESVDPDEL 60

DB 1 MSLSDWHLAVKLAADQPLTPKSIILRLPETELGYSLSGGYSISFLKQIAGKQESVDPDEL 60

QY 61 IDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSMPEPDQKPEVDKYAAMREFRVLHTAL 120

DB 61 IDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSMPEPDQKPEVDKYAAMREFRVLHTAL 120

QY 121 HSSSYREAVFVKMLSKESLQDIIVATPGLSSDPALGVLDKLFVSFADPNMGLTLPV 180

DB 121 HSSSYREAVFVKMLSKESLQDIIVATPGLSSDPALGVLDKLFVSFADPNMGLTLPV 180

QY 181 AHPALVNAIVLVLSVAGSAPMPGTQSSRSMPSSSYRDMPGGLFEGLSQDEDDHPNT 240

DB 181 AHPALVNAIVLVLSVAGSAPMPGTQSSRSMPSSSYRDMPGGLFEGLSQDEDDHPNT 240

QY 241 RSTPSSSTFSSRPASLGYSAGAGRPITQSELATALASTPSSSHTPTPGTQGHSSGT 300

DB 241 RSTPSSSTFSSRPASLGYSAGAGRPITQSELATALASTPSSSHTPTPGTQGHSSGT 300

QY 301 SPSSGVQSGTPTITNLFQQA-CHALQALQASQPSLSQSQWQFQQLQRLDMGICDDLSLRPC 360

DB 301 SPSSGVQSGTPTITNLFQQA-CHALQALQASQPSLSQSQWQFQQLQRLDMGICDDLSLRPC 360

QY 361 RPPVGTSKQWSSSLLEPHELPAPEPPASCRGCPWEALMKVPPSLP 409

DB 361 RPPVGTSKQWSSSLLEPHELPAPEPPASCRGCPWEALMKVPPSLP 409

RESULT 2

AA15413

IJ AA15413 standard; Protein; 426 AA.

XX AA15413;

CT 06-JUL-1997 (first entry)

XX Human activated platelet protein 2 APP 2.

XX APP 2; activated platelet protein-2; (thrombus; monoclonal antibody

XX Homo sapiens.

XX WO9712898-A1.

XX 10-APR-1997.

XX 04-OCT-1996; 96WO-US15922.

XX 06-OCT-1995; 95US-0005074.

XX (HARC) HARVARD COLLEGE.

XX Clement CY, Reed G;

XX WPI; 1997-226153/20.

DR N-PSDB; AA164549

XX Naturally-occurring platelet activation polypeptide, APP-2 - used

PT for localising a platelet thrombus in an animal, or activated platelet

PT complex in a biological sample

XX Claim 1; Fig 5; 50pp; English.

PS A novel 25 kDa polypeptide (AA15413) designated activated platelet

CC protein-2 (APP-2) is preferentially expressed on activated human

CC platelets but not resting platelets. APP-2 is characterised as

CC confg. an epitope that binds to monoclonal antibody 3B2 (ATCC CRL

CC 11986). Its amino acid sequence was deduced from a cDNA clone

CC (AA164549) obtd. from a human bone marrow cDNA library. An

CC alternatively spliced variant (see also AA15414) was also isolated.

CC APP-2 can be produced in transformed host cells. Anti-APP-2

CC antibodies are useful in methods for detecting activated platelets

CC in biological samples, for localising a platelet thrombus in an

CC animal, and for detecting an activated platelet protein complex in

CC an animal.

XX SQ Sequence 426 AA;

Query Match 99.9%; Score 2129; DB 18; Length 426;

Best Local Similarity 100.0%; Pred. No. 5.1e-167;

Matches 409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLSDWHLAVKLAADQPLTPKSIILRLPETELGYSLSGGYSISFLKQIAGKQESVDPDEL 60

DB 1 MSLSDWHLAVKLAADQPLTPKSIILRLPETELGYSLSGGYSISFLKQIAGKQESVDPDEL 60

QY 61 IDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSMPEPDQKPEVDKYAAMREFRVLHTAL 120

DB 61 IDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSMPEPDQKPEVDKYAAMREFRVLHTAL 120

QY 121 HSSSYREAVFVKMLSKESLQDIIVATPGLSSDPALGVLDKLFVSFADPNMGLTLPV 180

DB 121 HSSSYREAVFVKMLSKESLQDIIVATPGLSSDPALGVLDKLFVSFADPNMGLTLPV 180

QY 181 AHPALVNAIVLVLSVAGSAPMPGTQSSRSMPSSSYRDMPGGLFEGLSQDEDDHPNT 240

DB 181 AHPALVNAIVLVLSVAGSAPMPGTQSSRSMPSSSYRDMPGGLFEGLSQDEDDHPNT 240

QY 241 RSTPSSSTFSSRPASLGYSAGAGRPITQSELATALASTPSSSHTPTPGTQGHSSGT 300

DB 241 RSTPSSSTFSSRPASLGYSAGAGRPITQSELATALASTPSSSHTPTPGTQGHSSGT 300

QY 301 SPSSGVQSGTPTITNLFQQA-CHALQALQASQPSLSQSQWQFQQLQRLDMGICDDLSLRPC 360

DB 301 SPSSGVQSGTPTITNLFQQA-CHALQALQASQPSLSQSQWQFQQLQRLDMGICDDLSLRPC 360

QY 361 RPPVGTSKQWSSSLLEPHELPAPEPPASCRGCPWEALMKVPPSLP 409

DB 361 RPPVGTSKQWSSSLLEPHELPAPEPPASCRGCPWEALMKVPPSLP 409

RESULT 3

AA15413

ID AA15413 standard; Protein; 420 AA.

XX AA15413;

XX 09-DEC-1999 (first entry)

XX Amino acid sequence of fx317_11.

XX secreted protein; cDNA library; clone; transmembrane protein;

XX signal sequence cloning; hybridization cloning; gene therapy;

XX receptor.

02-JAN-1998; 98US-070346P.
07-JAN-1998; 98US-070643P.
08-JAN-1998; 98US-070755P.
13-JAN-1998; 98US-071134P.
22-JAN-1998; 98US-072134P.
30-JAN-1998; 98US-073095P.
18-FEB-1998; 98US-075218P.
30-MAR-2000; 2000US-2539130.
23-NOV-1998; 98US-1197864.
(JACO//) JACOBS K.
(MCCO//) MCCOY C. M.
(LAVA//) LAVALLIE E. R.
(COLL//) COLLINS RACIE L. A.
(EVAN//) EVANS C.
(MERB//) MERBERG D.
(TEEA//) TREACY M.
(ACOS//) ACOSTINO V. J.
(STEI//) STEININGER P. J.
(SPAU//) SPAULDING V.
(WONG//) WONG G. G.
(CLAR//) CLARK H.
(FECU//) FECHTEL K.
Jacobs K, McCoy C. M., Lavallie E. R., Collins-Racie L. A., Evans C.,
Merberg D., Tracy M., Agostino M. J., Steininger P. J., Spaulding V.,
Wong G. G., Clark H., Fechtel K;
WP1; 2002-040735/05.
NP1; 2002-040735/05.
NP2; 2002-040735/05.
NP3; 2002-040735/05.
NP4; 2002-040735/05.
NP5; 2002-040735/05.
NP6; 2002-040735/05.
NP7; 2002-040735/05.
NP8; 2002-040735/05.
NP9; 2002-040735/05.
NP10; 2002-040735/05.
NP11; 2002-040735/05.
NP12; 2002-040735/05.
NP13; 2002-040735/05.
NP14; 2002-040735/05.
NP15; 2002-040735/05.
NP16; 2002-040735/05.
NP17; 2002-040735/05.
NP18; 2002-040735/05.
NP19; 2002-040735/05.
NP20; 2002-040735/05.
NP21; 2002-040735/05.
NP22; 2002-040735/05.
NP23; 2002-040735/05.
NP24; 2002-040735/05.
NP25; 2002-040735/05.
NP26; 2002-040735/05.
NP27; 2002-040735/05.
NP28; 2002-040735/05.
NP29; 2002-040735/05.
NP30; 2002-040735/05.
NP31; 2002-040735/05.
NP32; 2002-040735/05.
NP33; 2002-040735/05.
NP34; 2002-040735/05.
NP35; 2002-040735/05.
NP36; 2002-040735/05.
NP37; 2002-040735/05.
NP38; 2002-040735/05.
NP39; 2002-040735/05.
NP40; 2002-040735/05.
NP41; 2002-040735/05.
NP42; 2002-040735/05.
NP43; 2002-040735/05.
NP44; 2002-040735/05.
NP45; 2002-040735/05.
NP46; 2002-040735/05.
NP47; 2002-040735/05.
NP48; 2002-040735/05.
NP49; 2002-040735/05.
NP50; 2002-040735/05.
NP51; 2002-040735/05.
NP52; 2002-040735/05.
NP53; 2002-040735/05.
NP54; 2002-040735/05.
NP55; 2002-040735/05.
NP56; 2002-040735/05.
NP57; 2002-040735/05.
NP58; 2002-040735/05.
NP59; 2002-040735/05.
NP60; 2002-040735/05.
NP61; 2002-040735/05.
NP62; 2002-040735/05.
NP63; 2002-040735/05.
NP64; 2002-040735/05.
NP65; 2002-040735/05.
NP66; 2002-040735/05.
NP67; 2002-040735/05.
NP68; 2002-040735/05.
NP69; 2002-040735/05.
NP70; 2002-040735/05.
NP71; 2002-040735/05.
NP72; 2002-040735/05.
NP73; 2002-040735/05.
NP74; 2002-040735/05.
NP75; 2002-040735/05.
NP76; 2002-040735/05.
NP77; 2002-040735/05.
NP78; 2002-040735/05.
NP79; 2002-040735/05.
NP80; 2002-040735/05.
NP81; 2002-040735/05.
NP82; 2002-040735/05.
NP83; 2002-040735/05.
NP84; 2002-040735/05.
NP85; 2002-040735/05.
NP86; 2002-040735/05.
NP87; 2002-040735/05.
NP88; 2002-040735/05.
NP89; 2002-040735/05.
NP90; 2002-040735/05.
NP91; 2002-040735/05.
NP92; 2002-040735/05.
NP93; 2002-040735/05.
NP94; 2002-040735/05.
NP95; 2002-040735/05.
NP96; 2002-040735/05.
NP97; 2002-040735/05.
NP98; 2002-040735/05.
NP99; 2002-040735/05.
NP100; 2002-040735/05.
NP101; 2002-040735/05.
NP102; 2002-040735/05.
NP103; 2002-040735/05.
NP104; 2002-040735/05.
NP105; 2002-040735/05.
NP106; 2002-040735/05.
NP107; 2002-040735/05.
NP108; 2002-040735/05.
NP109; 2002-040735/05.
NP110; 2002-040735/05.
NP111; 2002-040735/05.
NP112; 2002-040735/05.
NP113; 2002-040735/05.
NP114; 2002-040735/05.
NP115; 2002-040735/05.
NP116; 2002-040735/05.
NP117; 2002-040735/05.
NP118; 2002-040735/05.
NP119; 2002-040735/05.
NP120; 2002-040735/05.
NP121; 2002-040735/05.
NP122; 2002-040735/05.
NP123; 2002-040735/05.
NP124; 2002-040735/05.
NP125; 2002-040735/05.
NP126; 2002-040735/05.
NP127; 2002-040735/05.
NP128; 2002-040735/05.
NP129; 2002-040735/05.
NP130; 2002-040735/05.
NP131; 2002-040735/05.
NP132; 2002-040735/05.
NP133; 2002-040735/05.
NP134; 2002-040735/05.
NP135; 2002-040735/05.
NP136; 2002-040735/05.
NP137; 2002-040735/05.
NP138; 2002-040735/05.
NP139; 2002-040735/05.
NP140; 2002-040735/05.
NP141; 2002-040735/05.
NP142; 2002-040735/05.
NP143; 2002-040735/05.
NP144; 2002-040735/05.
NP145; 2002-040735/05.
NP146; 2002-040735/05.
NP147; 2002-040735/05.
NP148; 2002-040735/05.
NP149; 2002-040735/05.
NP150; 2002-040735/05.
NP151; 2002-040735/05.
NP152; 2002-040735/05.
NP153; 2002-040735/05.
NP154; 2002-040735/05.
NP155; 2002-040735/05.
NP156; 2002-040735/05.
NP157; 2002-040735/05.
NP158; 2002-040735/05.
NP159; 2002-040735/05.
NP160; 2002-040735/05.
NP161; 2002-040735/05.
NP162; 200

QY 61 IDLIYCGKXKDDQDLPFGTQPGSTVHVLKSNPEPCCKPEPVCKVAAAREFRVLTAL 120
 DB 101 IDLIYCGKXKDDQDLPFGTQPGSTVHVLKSNPEPCCKPEPVCKVAAAREFRVLTAL 160
 QY 121 HSSSYRAVAFKMLSNKESLDQIIVATPGLSQALQALQASGQPSLOSQWQFCQDLRMGICDDELSLR 180
 DB 161 HSSSYRAVAFKMLSNKESLDQIIVATPGLSQALQALQASGQPSLOSQWQFCQDLRMGICDDELSLR 220
 QY 181 AHPALVNAIVLVHVSAGSAMPQTDSSSRMPSSSYRDMFGFLFEGLSDDDDFHNT 240
 DB 221 AHPALVNAIVLVHVSAGSAMPQTDSSSRMPSSSYRDMFGFLFEGLSDDDDFHNT 280
 QY 241 RSTPSSSTPSSRPASLGYSAGGAPRPTQSELATALALASTPSSSHTPTPTGCHSSGT 300
 DB 281 RSTPSSSTPSSRPASLGYSAGGAPRPTQSELATALALASTPSSSHTPTPTGCHSSGT 340
 QY 301 SPMSSGVOSGTPITNCLFSQALQALQASGQPSLOSQWQFCQDLRMGICDDELSLR 358
 DB 341 SPMSSGVOSGTPITNCLFSQALQALQASGQPSLOSQWQFCQDLRMGICDDELSLR 396

RESULT 6
 AAG33292
 ID AAG33292 standard; Protein; 380 AA.
 AC AAG33292;
 XX
 DT 13-SEP-2001 (first entry;
 XX Human protein: HP10657.
 DE Human; gene therapy; tumour.
 KW Homo sapiens.
 OS Homo sapiens.
 XX
 FN WO200242302-A1.
 PC 14-JUN-2001.
 XX
 FE 06-DEC-2000; 2000ACJP08631.
 XX
 PF 06-DEC-1999; 99JP-0346663.
 PF 06-DEC-1999; 99JP-0346664.
 PF 08-FEB-2000; 2000JP-0031062.
 PF 10-FEB-2000; 2000JP-0034090.
 PF 10-FEB-2000; 2000JP-0034091.
 PF 14-FEB-2000; 2000JP-0035829.
 PF 14-FEB-2000; 2000JP-0035849.
 PF 14-MAR-2000; 2000JP-0031161.
 PF 30-MAY-2000; 2000JP-14-054.
 XX
 PA (NISC) JAPAN SCI & TECHNOLOGY
 PI Kato S, Eguchi C, Saeki M,
 XX
 XX WPI; 2001-391646/40.
 DR N-PSDB; AAH6R577.
 XX
 XX Human protein originated from tumour cell line, applicable as drug,
 PT reagent for studying intracellular protein networks and protein source
 PT for drug screening, also encoded data for gene diagnosis and gene
 PT therapy.
 XX
 XX Claim 1; Pages 313-314; 47pp; Japanese.
 PS
 CC The present sequence is a human protein. The human protein, preferably
 CC originated from tumour cell line, is applicable as a drug, a reagent for
 CC studying intracellular protein networks and a protein source for
 CC screening proteins for binding low molecular weight drugs. The human
 CC protein coding sequence is useful for gene diagnosis and gene therapy.
 CC expression vectors and transformant cells for detection of ligands and
 CC receptors.

XX SQ Sequence 380 AA;
 Query Match 85.7%; Score 1828; DB 22; Length 380;
 Best Local Similarity 99.7%; Pred. No. 2.9e-142;
 Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSLSNWHAVKLAQCPKLSILRPETELCEYSLGGYSISFLKQLIAGKLCSSVPDPPEL 60
 DB 1 MSLSNWHAVKLAQCPKLSILRPETELCEYSLGGYSISFLKQLIAGKLCSSVPDPPEL 60
 QY 61 IDLIYCGKXKDDQDLPFGTQPGSTVHVLKSNPEPCCKPEPVCKVAAAREFRVLTAL 120
 DB 61 IDLIYCGKXKDDQDLPFGTQPGSTVHVLKSNPEPCCKPEPVCKVAAAREFRVLTAL 120
 QY 121 HSSSYRAVAFKMLSNKESLDQIIVATPGLSQALQALQASGQPSLOSQWQFCQDLRMGICDDELSLR 180
 DB 121 HSSSYRAVAFKMLSNKESLDQIIVATPGLSQALQALQASGQPSLOSQWQFCQDLRMGICDDELSLR 180
 QY 181 AHPALVNAIVLVHVSAGSAMPQTDSSSRMPSSSYRDMFGFLFEGLSDDDDFHNT 240
 DB 181 AHPALVNAIVLVHVSAGSAMPQTDSSSRMPSSSYRDMFGFLFEGLSDDDDFHNT 240
 QY 241 RSTPSSSTPSSRPASLGYSAGGAPRPTQSELATALALASTPSSSHTPTPTGCHSSGT 300
 DB 241 RSTPSSSTPSSRPASLGYSAGGAPRPTQSELATALALASTPSSSHTPTPTGCHSSGT 300
 QY 301 SPMSSGVOSGTPITNCLFSQALQALQASGQPSLOSQWQFCQDLRMGICDDELSLR 358
 DB 301 SPMSSGVOSGTPITNCLFSQALQALQASGQPSLOSQWQFCQDLRMGICDDELSLR 358

RESULT 7
 ABB89944
 ID ABB89944 standard; Protein; 341 AA.
 XX
 AC ABB89944;
 DT 24-MAY-2002 (first entry;
 XX Human polypeptide SEQ ID NO 2320.
 DE
 XX
 KW Cystostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 XX WO200190304-A1
 FN
 XX
 XX 29-NOV-2001.
 XX
 XX 18-MAY-2001; 2001WO-US16440.
 XX
 XX 19-MAY-2000; 2000US-205515P.
 PR
 XX (HUMAN) HUMAN GEMMAVE SCI INC.
 PA
 XX
 PI Birse CE, Rosen CA;
 XX
 XX WPI; 2002-122026/16.
 DR
 DR N-PSDB; ABL9C353.
 XX
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders.
 XX
 PS Claim 11; SEQ ID NO 2320; 2081pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABL89449-ABL90453) and proteins

(ABB89040-ABB9044) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and anti-agonists are useful in the diagnosis, treatment and prevention of (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral amyloid and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in an alternative form, directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 341 AA:

Query Match 79.0%; Score 1686; DB 23; Length 341;
 Best Local Similarity 98.2%; Pred. No. 1.1e-135;
 Matches 330; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSLSQMLAVKLAQGLPKSGLPEPELGLVSLGYSISFLKGLIAKLIQESVDFEL 60
 DB : MSLSQHLAVKLAQGLPKSGLPEPELGLVSLGYSISFLKGLIAKLIQESVDFEL 60
 QY 61 IDLVYGRKLDQDTDFYGIQSTVHVLRFSNPEPQKREYDKVAAYREFFVRLTAL 120
 DB : IDLVYGRKLDQDTDFYGIQSTVHVLRFSNPEPQKREYDKVAAYREFFVRLTAL 120
 QY 121 HSSSYRAVAFVYLNKESLDQIVATGSLNPAALGVKQSLFVSFAELNMLDTLP 180
 DB : HSSSYRAVAFVYLNKESLDQIVATGSLNPAALGVKQSLFVSFAELNMLDTLP 180
 QY 121 HSSSYRAVAFVYLNKESLDQIVATGSLNPAALGVKQSLFVSFAELNMLDTLP 180
 DB : HSSSYRAVAFVYLNKESLDQIVATGSLNPAALGVKQSLFVSFAELNMLDTLP 180
 QY 181 AAPALVNAIVLVHVSAGAPMEYSSSRVPSSEYDMEYTFDFEGLMDEGPHNT 240
 DB : AAPALVNAIVLVHVSAGAPMEYSSSRVPSSEYDMEYTFDFEGLMDEGPHNT 240
 QY 241 RPTPSSSTPSSRPASIGVGAARFETQSEATATALASTPSSSTPSSSTPSSSTP 300
 DB : RPTPSSSTPSSRPASIGVGAARFETQSEATATALASTPSSSTPSSSTPSSSTP 300
 QY 301 SPYSSVQSGTPTINDLFSCALCHLCAKALPQUS 316
 DB : SPYSSVQSGTPTINDLFSCALCHLCAKALPQUS 316
 QY 301 SPYSSVQSGTPTINDLFSCALCHLCAKALPQUS 316
 DB : SPYSSVQSGTPTINDLFSCALCHLCAKALPQUS 316

RESULT 9
 AAB63948
 ID AAB63948 standard; Protein: 164 AA
 XX AAB63948
 AC AAB63948
 XX AAB63948
 DT 26-MAR-2001 (first entry)
 DE Human prostate cancer associated antigen protein sequence SEQ ID NO:1310.
 XX Human prostate cancer associated antigen protein sequence SEQ ID NO:1311.
 KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
 KW cancer associated antigen; cytostatic; cancer vaccine.
 OS Homo sapiens.
 XX Homo sapiens.
 XX WC200073801-A2.
 XX WC200073801-A2.
 PD 07-DEC-2000.
 XX 07-DEC-2000.
 XX 26-MAY-2000; 2000WO-US14749.
 XX 26-MAY-1999; 99US-0136526.
 XX 10-SEP-1999; 99US-0153454.
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX

Query Match 37.2%; Score 794; DB 22; Length 191;
 Best Local Similarity 84.2%; Pred. No. 2.1e-57;
 Matches 165; Conservative 2; Mismatches 15; Indels 14; Gaps 2;

QY 96 EPQCKPEPVKVAAMREFRVJHTALHSSSYREAVFKMLSNKESLDQIVATPGLSSDPI 155
 DB : EPQCKPEPVKVAAMREFRVJHTALHSSSYREAVFKMLSNKESLDQIVATPGLSSDPI 155
 QY 156 ALGVLOKGLFVSFAELNMLDTLPAPALVNAIVLVHVSAGAPMPCTDSSSRMPSS 215
 DB : ALGVLOKGLFVSFAELNMLDTLPAPALVNAIVLVHVSAGAPMPCTDSSSRMPSS 215
 QY 216 SYRCPMGQSLFEGLSDEDECFHNTBSTPSSSTPSSRPASIGVGAARFETQSEATA 275
 DB : SYRCPMGQSLFEGLSDEDECFHNTBSTPSSSTPSSRPASIGVGAARFETQSEATA 275
 QY 276 LALASTPSSSTPSSRPASIGVGAARFETQSEATA 275
 DB : LALASTPSSSTPSSRPASIGVGAARFETQSEATA 275
 QY 167 LGPQSPFVSPPPPWP 182
 DB : LGPQSPFVSPPPPWP 182

PA (LUDW-) LUDWIG INST CANCER RES.
 XX Obata Y;
 XX WPI; 2001-025274/21;
 DR Nucleic acids encoding breast, gastric and prostate cancer associated
 PT antigen precursors, useful for diagnosing and treating a condition
 PT characterized by expression of an abnormal amount of a protein, e.g.
 PT cancer
 XX Example 1; Page 1-11, 799pp; English.
 XX AAF22422 to AAF22426, AAF22427 to AAF22773 and AAF22774 to AAF23014
 CC represent nucleic acid sequences encoding human breast, gastric and
 CC prostate cancer associated antigen precursors (CAAP) respectively.
 CC AAB63232 to AAB63447, AAB63468 to AAB63721 and AAB63722 to AAB63970
 CC represent human breast, gastric and prostate CAAP protein sequence
 CC respectively. CAAPs have cytostatic activity and can be used in the
 CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
 CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
 CC condition characterized by expression of an abnormal amount of a protein,
 CC e.g. cancer.
 XX SQ Sequence 191 AA:

```

PI Qata Y;
XX WPI; 2001-025274/03.
XX Nucleic acids encoding breast, gastric and prostate cancer associated
XX antigen precursors, useful for diagnosing and treating a condition
XX characterized by expression of an abnormal amount of a protein, e.g.
XX cancer.
XX Example 1; Page 789; 730pp; English.
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23114
XX represent nucleotide sequences encoding human breast, gastric and
XX prostate cancer associated antigen precursors (AA221, respectively,
XX AAF2312 to AAF23467, AAF23468 to AAF23721 and AAF23722 to AAF23799
XX represent human breast, gastric and prostate CAAP protein sequence
XX respectively. CAAPs have cytostatic activity and can be used in the
XX production of cancer vaccines. The human CAAP proteins, peptides, nucleic
XX acids or anti-CAAP antibodies are useful for diagnosing and treating a
XX condition characterised by expression of an abnormal amount of a protein,
XX e.g. cancer.
XX Sequence 164 AA;
SQ
Query Match 27.8%; Score 932; 24 22; Length 164;
Best Local Similarity 83.8%; Pred No. 7.6e-41;
Matches 119; Conservative 1; Mismatches 22; Indels 0; Gaps 0;
QY 217 YRMFGGFFEGLEDDEDFHPTNRTTSSSTTMRPSSGYSMADEPRPTTUSELATN 204
DB 1 YRDMFGASCLKGSMRMTFTQTPHFXSTPUSREASLGYSAGGPRPTUSELATAD 40
QY 277 ALASTPESSSHPTFGTGHSSGTSMSSGYGSGTPTNDELSALQHAALAGCGTSLOS 316
DB 61 ALASTPESSSHPTFGTGHSSGTSPKXGVSGSTPTNDELSALQHAALAGCGTSLOS 120
QY 337 QWQQLQQRDMGICQDELSLR 358
DB 121 QWQQLQQRDMGICQDELSLR 140
RESULT 10
ID AAG13489 standard; Protein; 128 AA
AC AAG13489;
XX
XX 06-NOV-2000; (first entry)
XX Human polypeptide SEQ ID NO: 13489.
XX Human; cytokine; cell proliferation, cell differentiation, gene therapy,
XX vaccine; peptide therapy; stem cell growth factor; hematopoiesis,
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation.
XX Homo sapiens.
XX WO200164835-A2.
XX 07-SEP-2001.
XX 26-FEB-2001; 2001WO-US04927.
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX (HYSE-); HYSEQ INC.
XX Tang YF, Liu C, Drmanac RT;
XX WPI; 2001-514839/56.
XX N PSCB; AAI93423.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders.
XX Claim 20; SEQ ID NO 27381; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AA000330-AA013910) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, hematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX Sequence 128 AA;
SQ
Query Match 26.5%; Score 566; DB 22; Length 128;
Best Local Similarity 96.4%; Pred No. 7.6e-39;
Matches 108; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSLSDWHLAVNADQPLTPKSLRLPTELTGEVSLGYSISFUKQLIAGKLOESVDPDEL 60
DB 17 MSLSDWHLAAKLADQPLTPKSLRLPTELTGEVSLGYSISFUKQLIAGKLOESVDPDEL 76
QY 61 IDLLYCGRKLLKDDCTLDYGLQPGSTVHVLKSWPEPDQKPEPVDKVAAMRE 112
DB 77 IDLLYCGRKLLKDDCTLDYGLQPGSTVHVLKSWPEPDQKPEPVDKVAAMRD 128
RESULT 11;
ID AAG02759 standard; Protein; 76 AA.
XX AAG02759;
AC AAG02759;
XX
XX 06-OCT-2000 (first entry)
XX Human secreted protein, SEQ ID NO: 6840.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping.
XX Homo sapiens.
XX EP033401-A2.
XX C6-SEP-2000.
XX 21-FEB-2000; 2000EP-0200610.
XX 26-FEB-1999; 99US-0122487.
XX (GEST ) GENSET.
XX Dunas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX N-PSCB; AAG02765.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 13; SEQ ID 6840; 71pp + CD-ROM; English.
XX

```



```
PR 21 OCT-1999; 99US-0160768
PR 21 OCT-1999; 99US-0160770
PR 21 OCT-1999; 99US-0160814
PR 21 OCT-1999; 99US-0160815
PR 22 OCT-1999; 99US-0160980
PR 22 OCT-1999; 99US-0160981
PR 22 OCT-1999; 99US-0160989
PR 22 OCT-1999; 99US-0161404
PR 22 OCT-1999; 99US-0161405
PR 25 OCT-1999; 99US-0161406
PR 25 OCT-1999; 99US-0161359
PR 26 OCT-1999; 99US-0161360
PR 26 OCT-1999; 99US-0161361
PR 28 OCT-1999; 99US-0161320
PR 28 OCT-1999; 99US-0161392
PR 28 OCT-1999; 99US-0161393
PR 29 OCT-1999; 99US-0162142

Query Match 7.7% Score 164.5 EPR 21; length 536.
Best Local Similarity 24.0%; P-Val NO. 7.2e-05;
Matches 99; Conservative 49; Matches 140; Index 125; gaps 19;

QY 39 SISFLKQIAGKQESVPEPT:GVYKPK:KQCTIDFKGIQPSWIKWKWEPD 98
D- |||||
D- |||||
D- |||||
QY 39 TVSEFKELIA---QNSDFPAGCPHYKGR::KQCTLLSVNLOADHTVWYVQVPS 95
QY 99 QAP-----EPVNVAAVREF .....RV 115
D- ||
D- ||
D- ||
QY 96 SAPAAAGNQTTPACVGSNDSSVLOGGSLFPGSPNPGSNVAVGEPFSGLPCLLA 155
QY 116 LHAIDSSSYE ..AVKMKKESLDTIVATKASQPIALGVLOOK DLSV 168
D- |||||
D- |||||
D- |||||
QY 156 QCLAQKPMREKMTPLAKNNPFRAMNNPNKRE-----DVRNDEL RV 208
QY 169 FAPNMC DTLVPA HPALVNAIIVHSAVHAIWGTDSNMSKSYKPMFSG 224
D- |||||
D- |||||
D- |||||
QY 209 LNPRLRLCTLEARNPELMREV ...NRND RAM-SNLSMSEG-NM 251
QY 225 --LFEGLSDDDDEHENTSTIGTFSSEF ASQHSWAGRPPTGCLLAAGLA 279
D- |||||
D- |||||
D- |||||
QY 252 LRMVENV--QPLAKNTTSENAGNNTSSMFAALGNGNTTQOSTSNKISAPRET 309
QY 280 STTESSEHTPTQSHSG--LHMWSTVQ... ..ETRN 315
D- |||||
D- |||||
D- |||||
QY 316 LNPRLRLCTLEARNPELMREV ...NRND RAM-SNLSMSEG-NM 251
D- |||||
D- |||||
D- |||||
QY 367 QL SCLQNFVSKMGNVLSNIPVSGNNGNQLKMMVPLRKNY... 414

RESULT 14
AAG30039
ID AAG30039 standard; Protein; 577 AA.
XX
AC AAG30039;
XX
XX 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SE; ID NO. 1-843.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
CS Arabidopsis thaliana.
XX
PN EP10334C5-A2.
XX
FC 06-SEP-2000.
XX
IF 25-FEB-2000, 2:00EP-030139.
XX
```

```
PR 13-JUL-1999; 99US-0143542;
PR 14-JUL-1999; 99US-0143624;
PR 15-JUL-1999; 99US-0144025;
PR 16-JUL-1999; 99US-0144285;
PR 16-JUL-1999; 99US-0144386;
PR 19-JUL-1999; 99US-0144325;
PR 19-JUL-1999; 99US-0144331;
PR 19-JUL-1999; 99US-0144332;
PR 19-JUL-1999; 99US-0144333;
PR 19-JUL-1999; 99US-0144334;
PR 19-JUL-1999; 99US-0144335;
PR 20-JUL-1999; 99US-0144352;
PR 20-JUL-1999; 99US-0144632;
PR 20-JUL-1999; 99US-0144684;
PR 21-JUL-1999; 99US-0144614;
PR 21-JUL-1999; 99US-0145065;
PR 21-JUL-1999; 99US-0145088;
PR 22-JUL-1999; 99US-0145085;
PR 22-JUL-1999; 99US-0145087;
PR 22-JUL-1999; 99US-0145089;
PR 22-JUL-1999; 99US-0145192;
PR 23-JUL-1999; 99US-0145145;
PR 23-JUL-1999; 99US-0145218;
PR 23-JUL-1999; 99US-0145224;
PR 26-JUL-1999; 99US-0145276;
PR 27-JUL-1999; 99US-0145913;
PR 27-JUL-1999; 99US-0145918;
PR 27-JUL-1999; 99US-0145919;
PR 28-JUL-1999; 99US-0145951;
PR 28-AUG-1999; 99US-0146166;
PR 28-AUG-1999; 99US-0146189;
PR 28-AUG-1999; 99US-0146389;
PR 03-AUG-1999; 99US-0147018;
PR 04-AUG-1999; 99US-0147224;
PR 05-AUG-1999; 99US-0147322;
PR 05-AUG-1999; 99US-0147192;
PR 06-AUG-1999; 99US-0147260;
PR 06-AUG-1999; 99US-0147303;
PR 06-AUG-1999; 99US-0147416;
PR 09-AUG-1999; 99US-0147493;
PR 09-AUG-1999; 99US-0147935;
PR 10-AUG-1999; 99US-0148171;
PR 11-AUG-1999; 99US-0148319;
PR 12-AUG-1999; 99US-0148341;
PR 13-AUG-1999; 99US-0148565;
PR 13-AUG-1999; 99US-0148684;
PR 16-AUG-1999; 99US-0149368;
PR 17-AUG-1999; 99US-0149115;
PR 18-AUG-1999; 99US-0149425;
PR 20-AUG-1999; 99US-0149422;
PR 20-AUG-1999; 99US-0149513;
PR 20-AUG-1999; 99US-0149929;
PR 23-AUG-1999; 99US-0149902;
PR 23-AUG-1999; 99US-0149930;
PR 25-AUG-1999; 99US-0150566;
PR 26-AUG-1999; 99US-0150884;
PR 27-AUG-1999; 99US-0151065;
PR 27-AUG-1999; 99US-0151066;
PR 27-AUG-1999; 99US-0151080;
PR 30-AUG-1999; 99US-0151303;
PR 31-AUG-1999; 99US-0151438;
PR 01-SEP-1999; 99US-0151935;
PR 17-SEP-1999; 99US-0152363;
PR 10-SEP-1999; 99US-0153070;
PR 13-SEP-1999; 99US-0153758;
PR 15-SEP-1999; 99US-0154018;
PR 16-SEP-1999; 99US-0154039;
PR 20-SEP-1999; 99US-0154779;
PR 22-SEP-1999; 99US-0155139;
PR 23-SEP-1999; 99US-0155486;
PR 24-SEP-1999; 99US-0155659;
PR 28-SEP-1999; 99US-0156458;
PR 29-SEP-1999; 99US-0156596;

PR 04-OCT-1999; 99US-0157117;
PR 05-OCT-1999; 99US-0157753;
PR 06-OCT-1999; 99US-0157865;
PR 07-OCT-1999; 99US-0158029;
PR 08-OCT-1999; 99US-0158232;
PR 12-OCT-1999; 99US-0158369;
PR 13-OCT-1999; 99US-0159293;
PR 13-OCT-1999; 99US-0159294;
PR 13-OCT-1999; 99US-0159295;
PR 14-OCT-1999; 99US-0159329;
PR 14-OCT-1999; 99US-0159330;
PR 14-OCT-1999; 99US-0159331;
PR 14-OCT-1999; 99US-0159637;
PR 18-OCT-1999; 99US-0159638;
PR 18-OCT-1999; 99US-0159584;
PR 21-OCT-1999; 99US-0162741;
PR 21-OCT-1999; 99US-0162741;
PR 21-OCT-1999; 99US-0162767;
PR 21-OCT-1999; 99US-0162768;
PR 21-OCT-1999; 99US-0160770;
PR 21-OCT-1999; 99US-0160814;
PR 22-OCT-1999; 99US-0160815;
PR 22-OCT-1999; 99US-0160980;
PR 22-OCT-1999; 99US-0160981;
PR 22-OCT-1999; 99US-0160989;
PR 25-OCT-1999; 99US-0161404;
PR 25-OCT-1999; 99US-0161405;
PR 25-OCT-1999; 99US-0161406;
PR 26-OCT-1999; 99US-0161359;
PR 26-OCT-1999; 99US-0161360;
PR 26-OCT-1999; 99US-0161361;
PR 28-OCT-1999; 99US-0161920;
PR 28-OCT-1999; 99US-0161942;
PR 28-OCT-1999; 99US-0161993;
PR 29-OCT-1999; 99US-0162142;

Query Match 7.7%; Score 154.5; DB 21; Length 577;
Best Local Similarity 24.0%; Pred No Re-05;
Matches 99; Conservative 49; Mismatches 140; Indels 125; Gaps 19;

QY 33 S-SF-KC-IAGKLOESVPDPEDLIYCGRKLKDDQTLDFYGIQPGSTVHVLRKSWPEPD 98
DB 80 TVESFEKELIA--QNSDVPANQORLYKGRILKDDQTLISYGLQADHTVHMVGRGVPSPP 136
QY 99 QXP-----EPDKVAAMREF-----RV 115
DB 137 SAPAAVAGNQTITAPQAVGNSNDSSNUGGESLPGLGFNPLGGNAGACLFSGGLPDLQEA 196
QY 116 LHTALSSSSVRE----AVFKMLSNKESLDGCIIVATPGLSSDPALGVLOCK--DGFVS 168
DB 19 QOCLAJNFMIREMNTPAIGNKXNPEFYKSMINMPCRE-----JVRNPELGHV 249
QY 169 FADPNMLDTLTPA-HPALVNAIVLVHVSAGSAPYPTGTDSSSRSPSSSYRMPGGF-- 224
DB 250 LNDPSILQTLQTLGAANPELMREM-----MRNTD--RAM--SNIESMPEGNM 292
QY 225 ---LFEGLSDDEDDPHNTRSTPSSSTPSSRP--ASLCYSGAAGPRPITQSELATALLA 279
DB 293 LRRMYENV--QEPMLXATTTMGENAGNTSSNPFALLGNQGVTTQGSDTSNK--SAPNAET 350
QY 280 STPESSSHPTPTGTOGRSSG--TSPWSSGVQSG-----TPITN 315
DB 351 GTPNAN- PIPNPGATAGQTTAPGRTNAGLGGGLGGGLGGLGMLGADSPLAGTTPDAS 407
QY 316 DLFSQALOHALQASQPSQSOWQ-----POLQOQRDMG;QDBELSLRP 359
DB 408 QL-SQILCNAPMSQMWQSVLSNPQYMCNLSINPOLRSMIDMNPOLREMGNQP 459

RESULT 15
ABR69125
ID ABR69125 standard; Protein; 547 AA.
XX
AC ABR69125;
```


GerCore version 5.1.6
Copyright (c) 1993 2003 Corpuagen Ltd.

OM protein - protein search, using sw modli

Run on: November 12, 2003, 06:18:19 / Search time 30 Seconds
without alignments
234 240 Million cell: updates/sec

Title: US-09-735-251-4

Perfect score: 2133

Sequence: 1 MSLSMR:AVKLADQLTPK... SPPGYCWEADYKVSSEV 410

Scoring table: RLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications: AA*

1: /cgn2_6/ptodata/1/pubtaa/US01_PUBCOMB pep.*
2: /cgn2_6/ptodata/1/pubtaa/US01_PUBCOMB pep.*
3: /cgn2_6/ptodata/1/pubtaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubtaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubtaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubtaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubtaa/US08_NEW_PUB pep.*
8: /cgn2_6/ptodata/1/pubtaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubtaa/US09_PUBCOMB pep.*
10: /cgn2_6/ptodata/1/pubtaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubtaa/US09_PUBCOMB pep.*
12: /cgn2_6/ptodata/1/pubtaa/US09_NEW_PUB pep.*
13: /cgn2_6/ptodata/1/pubtaa/US10_PUBCOMB pep.*
14: /cgn2_6/ptodata/1/pubtaa/US10_PUBCOMB pep.*
15: /cgn2_6/ptodata/1/pubtaa/US10_PUBCOMB pep.*
16: /cgn2_6/ptodata/1/pubtaa/US10_PUBCOMB pep.*
17: /cgn2_6/ptodata/1/pubtaa/US10_PUBCOMB pep.*
18: /cgn2_6/ptodata/1/pubtaa/US10_PUBCOMB pep.*

Pred. No. is the number of predicted hits that have a score greater than or equal to the score of the hit being ranked, and is derived by analysis of the total database distribution

SUMMARY

Result No.	Score	Query Match	Length	ID	Description
1	1834	86.0	380	12	US-09-692-684-46 Sequence 86, Appl
2	1834	86.0	420	9	US-09-729-674-192 Sequence 172, Appl
3	152.5	7.1	124	15	US-10-102-946-438 Sequence 488, Appl
4	152.5	7.1	624	15	US-10-146-474-47 Sequence 47, Appl
5	141.5	6.6	589	12	US-10-293-000-4 Sequence 4, Appl
6	141.5	6.6	595	12	US-10-293-000-2 Sequence 2, Appl
7	141	6.6	528	12	US-09-840-746-20 Sequence 20, Appl
8	135.5	6.4	1162	10	US-09-745-083-34 Sequence 34, Appl
9	134	6.3	655	14	US-10-601-612A-2 Sequence 2, Appl
10	134	6.3	655	15	US-10-054-681-29 Sequence 29, Appl
11	128	6.0	1134	10	US-09-816-392-16 Sequence 16, Appl
12	127.5	6.0	881	9	US-09-816-860A-2 Sequence 2, Appl
13	126	5.9	2971	15	US-10-146-473-52 Sequence 50, Appl
14	124.5	5.8	841	12	US-10-029-306-320A Sequence 320A, A
15	124.5	5.8	1714	15	US-10-128-714-3176 Sequence 3176, Ap

Sequence 8176, Ap
Sequence 3, Appli
Sequence 8, Appli
Sequence 264, App
Sequence 6, Appli
Sequence 4, Appli
Sequence 2, Appli
Sequence 92, Appli
Sequence 194, App
Sequence 761, App
Sequence 2, Appli
Sequence 2, Appli
Sequence 11, App
Sequence 5, Appli
Sequence 32888, A
Sequence 108, Appl
Sequence 78, Appl
Sequence 56, Appli
Sequence 2, Appli
Sequence 814, App
Sequence 3473, Ap
Sequence 1138, Ap
Sequence 14009, A
Sequence 2, Appli
Sequence 632, App
Sequence 4, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-890-688-86
; Sequence 86, Application: US/09890688
; Publication No. US20030144475A1
; GENERAL INFORMATION:
; APPLICANT: Setahi KATO
; APPLICANT: Minoro SAKI
; TITLE OF INVENTION: Human Proteins and cDNAs thereof
; FILE REFERENCE: 2001-1102A/WMC/C0653
; CURRENT APPLICATION NUMBER: US/09/890.688
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: JP 11-346863
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: JP 11-34684
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: JP 2000-31062
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: JP 2000-34091
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: JP 2000-34090
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: JP 2000-35829
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 2000-35899
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 2000-71161
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: JP 2000-160851
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 86
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-890-688-86

Query Match 86.0%; Score 1834; DB 12; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1, 1e-134;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLSDWHLAVKLADQPLTPKSIILPLPETELGEVSLGGYSISFLKQIAGKLOESVDPDEL 60
 DB 1 MSLSDWHLAVKLADQPLTPKSIILPLPETELGEVSLGGYSISFLKQIAGKLOESVDPDEL 60

QY 61 IDLIYCGRKLKDDQTLDFYGCQSTVHVLRRKSNPEFDQKIFVQKVAAMREFRVLTAL 120
 DB 61 IDLIYCGRKLKDDQTLDFYGCQSTVHVLRRKSNPEFDQKIFVQKVAAMREFRVLTAL 120

QY 121 HSSSYREAVFKMLSNKESLDQIIIVATPGLSSDPPIALGVLDQKDLFSVFADPNMMLDTLVP 180
 DB 121 HSSSYREAVFKMLSNKESLDQIIIVATPGLSSDPPIALGVLDQKDLFSVFADPNMMLDTLVP 180

QY 181 AHPALVNAIVLVLSVAGSAPVPGTDSRSSRPSSSYRDMPCGFLFEGLSDDDDPHFNT 240
 DB 181 AHPALVNAIVLVLSVAGSAPVPGTDSRSSRPSSSYRDMPCGFLFEGLSDDDDPHFNT 240

QY 241 RSTPSSSTPSRRPASLGYSGAAGPRPTTQSELATATLALASTPSSSHTPTPTQTQGHSSGT 300
 DB 241 RSTPSSSTPSRRPASLGYSGAAGPRPTTQSELATATLALASTPSSSHTPTPTQTQGHSSGT 300

QY 301 SPSSSGVCGSTPTINDLFSQALQHALQASQFSLOSOWOPQQLQURDMGIOQDELSLR 358
 DB 301 SPSSSGVCGSTPTINDLFSQALQHALQASQFSLOSOWOPQQLQURDMGIOQDELSLR 358

RESULT 2

US-09-729-674-172
 ; Sequence 172, Application US/09/129674
 ; Publication No. US20010339335A;
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John M.
 ; APPLICANT: LaVallie, Edward R.
 ; APPLICANT: Collins-Racie, Lisa A.
 ; APPLICANT: Evans, Cheryl
 ; APPLICANT: Moberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Agostino, Michael J.
 ; APPLICANT: Steinginger II, Robert J.
 ; APPLICANT: Spaulding, Vikki
 ; APPLICANT: Wong, Gordon G.
 ; APPLICANT: Clark, Hilary
 ; APPLICANT: Fichtel, Kim
 ; APPLICANT: Genetics International
 ; TITLE OF INVENTION: RECOMBINANT PROTEIN
 ; FILE REFERENCE: 6055-64X
 ; CURRENT APPLICATION NUMBER: US/09/129674
 ; PRIOR FILING DATE: 2000-03-10
 ; PRIOR APPLICATION NUMBER: 09/553,410
 ; NUMBER OF SEQ ID NOS: 283
 ; SOFTWARE: PatentIn Ver. 2.3
 ; SEQ ID NO 172
 ; LENGTH: 420
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-729-674-172

Query Match 86.0%; Score 1834; DB 9; Length 410;
 Best Local Similarity 100.0%; Pred. No. 1, 1e-134;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLSDWHLAVKLADQPLTPKSIILPLPETELGEVSLGGYSISFLKQIAGKLOESVDPDEL 60
 DB 41 MSLSDWHLAVKLADQPLTPKSIILPLPETELGEVSLGGYSISFLKQIAGKLOESVDPDEL 100

QY 61 IDLIYCGRKLKDDQTLDFYGCQSTVHVLRRKSNPEFDQKIFVQKVAAMREFRVLTAL 120
 DB 101 IDLIYCGRKLKDDQTLDFYGCQSTVHVLRRKSNPEFDQKIFVQKVAAMREFRVLTAL 160

QY 121 HSSSYREAVFKMLSNKESLDQIIIVATPGLSSDPPIALGVLDQKDLFSVFADPNMMLDTLVP 180
 DB 161 HSSSYREAVFKMLSNKESLDQIIIVATPGLSSDPPIALGVLDQKDLFSVFADPNMMLDTLVP 220

QY 181 AHPALVNAIVLVLSVAGSAPVPGTDSRSSRPSSSYRDMPCGFLFEGLSDDDDPHFNT 240
 DB 221 AHPALVNAIVLVLSVAGSAPVPGTDSRSSRPSSSYRDMPCGFLFEGLSDDDDPHFNT 280

QY 241 RSTPSSSTPSRRPASLGYSGAAGPRPTTQSELATATLALASTPSSSHTPTPTQTQGHSSGT 300
 DB 281 RSTPSSSTPSRRPASLGYSGAAGPRPTTQSELATATLALASTPSSSHTPTPTQTQGHSSGT 340

QY 301 SPSSSGVCGSTPTINDLFSQALQHALQASQFSLOSOWOPQQLQURDMGIOQDELSLR 358
 DB 341 SPSSSGVCGSTPTINDLFSQALQHALQASQFSLOSOWOPQQLQURDMGIOQDELSLR 398

RESULT 3

US-10-102-806-488
 ; Sequence 488, Application US/10102806
 ; Publication No. US20030034421A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA/03/PC1
 ; CURRENT APPLICATION NUMBER: US/10/102,806
 ; CURRENT FILING DATE: 2002-03-22
 ; PRIOR APPLICATION NUMBER: 09/925,298
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05881
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 62/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 846
 ; SOFTWARE: PatentIn Ver. 2.3
 ; SEQ ID NO 488
 ; LENGTH: 114
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (95)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (111)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (111)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; US-10-102-806-488

Query Match 17.3%; Score 370; DB 15; Length 114;
 Best Local Similarity 94.7%; Pred. No. 2, 3e-21;
 Matches 72; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSLSDWHLAVKLADQPLTPKSIILPLPETELGEVSLGGYSISFLKQIAGKLOESVDPDEL 60
 DB 22 MSLSDWHLAVKLADQPLTPKSIILPLPETELGEVSLGGYSISFLKQIAGKLOESVDPDEL 81

QY 61 IDLIYCGRKLKDDQTL 76
 DB 82 IDLIYCGRKLKDDQTL 97

RESULT 4

US-10-146-473-47
 ; Sequence 47, Application US/10146473
 ; Publication No. US20030108888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Scanlan, Matthew
 ; APPLICANT: Gout, Ivan
 ; APPLICANT: Stockert, Elisabeth

APPLICANT: Gure, Ali
 APPLICANT: Chen, Yao-Tsang
 APPLICANT: Old, Lloyd
 TITLE OF INVENTION: Breast Cancer Antigens
 FILE REFERENCE: US 60/461/70130 (GVR)
 CURRENT APPLICATION NUMBER: US/10/293-000-4
 CURRENT FILING DATE: 2001-05-15
 PRIOR APPLICATION NUMBER: US 60/293-000-2
 PRIOR FILING DATE: 2001-05-15
 NUMBER OF SEQ ID NOS: 82
 SOFTWARE: Patent in version 3.2
 SEQ ID NO 47
 LENGTH: 624
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-293-000-4

Query Match 7.1%; Score 152.5; DB 15; Length 624;
 Best Local Similarity 19.8%; Pred. No. 0.022;
 Matches 97; Conservative 70; Mismatches 17; Indels 147; Gaps 19;

QY 13 ADQPLTKS: LRLPETELGEYSL-GGYS:SLFKQLIAGKLOESVDPDLIIYCG 67
 DB 24 AAAPAEKIKVTVKTPK-EKEEFAPVENSVOQFKESIKRFXSHT---DOLVLIAG 79
 QY 69 KKKKDDQTLDFYGIQPGSTVHLRKSWEPPDKPEPVKVAAMREFRVLRLTAHSSSSYR 127
 DB 80 ILKQDQTLSCQGIHGLTVHLVIXTKNRPD-----HSAQQTNTAGS 125
 QY 128 EAVFKMLSKESLQCIIVATP-----GLSS-----DPI 155
 DB 131 SNSTPSTNSNPFGLSGGLAGLSSGLNTTNFSELSQSQMQRLLSNPEM 185
 QY 156 ALGVLOKDLFSVFADPNMLDILVPAHPALVNA:VL---VLHSVAGSAMPDGTSSSRSM 212
 DB 191 VQMSLNPFLMQILMANPQV-QGILKNPESLKNPQLINPQVTELEIANPAMQENR 249
 QY 209 SRPSSSYRMPGGF-LFEGSLDELFDEFTSTSSSTISRPASLYSNAG--- 263
 DB 250 NQDLALSNLESIPGGYNALRPYTDIPEKMAVEQFGSH---ASVSSSSSGGIC 306
 QY 264 -----PPITQSELATLALASTESSHT-----P 289
 DB 307 PRTYERKPLNPWAPPATSS ATSTSTTSJGNSNSNNTGNTYLMANVAFES 365
 QY 290 TPGIQQ-----MADPIMNSGVAIETATLITQIALAGLIDG 317
 DB 366 TPGQQLIQCITENPGLIQLNAGVYKSNM-----LQGLQGL 425
 QY 338 WCPHICQLRNGIQQDELSLRPRCPVGTSGKOPWSSSLLEPHELP 383
 DB 426 MRPG-----PFLQCMQGLTSAVNHFA----- 461
 QY 338 WEALMKVPPEL 408
 DB 452 MQALMQIQQGL 462

RESULT 5
 US-10-293-000-4
 Sequence 4, Application US/10293000
 Publication No. US20030175278A1
 GENERAL INFORMATION:
 APPLICANT: Monteliro, Mervyn J.
 APPLICANT: Mah, Alex L.
 APPLICANT: Perry, George
 APPLICANT: Smith, Mark A.
 TITLE OF INVENTION: UBIQUILIN, A PRESENILIN INTERACTOR AND METHODS OF USING SAME
 FILE REFERENCE: 4115.175
 CURRENT APPLICATION NUMBER: US/10/293-000
 PRIOR APPLICATION NUMBER: 2003-04-01
 CURRENT FILING DATE: 2003-04-01
 PRIOR FILING DATE: 2001-11-13

NUMBER OF SEQ ID NOS: 8
 SOFTWARE: Patent in version 3.2
 SEQ ID NO 4
 LENGTH: 589
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-293-000-4

Query Match 6.6%; Score 141.5; DB 12; Length 589;
 Best Local Similarity 18.6%; Pred. No. 0.014;
 Matches 94; Conservative 81; Mismatches 169; Indels 161; Gaps 19;

QY 9 AVKADCPKLPKSLRPLPETELGEYSL-GGYS:SLFKQLIAGKLOESVDPDLIIYCG 67
 DB 27 AAASAEPRXIKVTVKTPK-EKEEFAPVENSVOQFKESIKRFXSHT---DOLVLIAG 82
 QY 68 RKKKDDQTLDFYGIQPGSTVHLRKSWEPPDKPEPVKVAAMREFRVLRLTAHSSSSYR 127
 DB 83 KILKDDQTLSCQGIHGLTVHLVIXTKNRPD-----HSAQQTNTAGS 125
 QY 128 EAVFKMLSKESLQCIIVATP-----GLSS-----DPI 155
 DB 126 NVTTSSTPKNSSTSSGATSNPFGLSGGLAGLSSGLNTTNFSELSQSQMQRLLSNPEM 185
 QY 156 ALGVLOKDLFSVFADPNMLDILVPAHPALVNA:VL---VLHSVAGSAMPDGTSSSRSM 212
 DB 196 MVOIXENPFGVMSLNPDLINPQLINPQVTELEIANPAMQENR 244
 QY 213 PS-----SSYRMPGGF-----LFEG-----LSDDEDFHNTSTSTSSSTP 249
 DB 245 PAMQEMMNQDRALSNLESIPGGYNALRPYTDIPEKMAVEQFGSH---ASVSSSSSGGIC 304
 QY 250 SRPSSSYRMPGGF-LFEGSLDELFDEFTSTSSSTISRPASLYSNAG--- 303
 DB 305 SGEQSFSTENRCPNPWAPPATSS ATSTSTTSJGNSNSNNTGNTYLMANVAFES 363
 QY 304 SSGVOS-----GTP:TNDFLS-----CALQHALQA----- 328
 DB 364 VPGVQASMTNPGMOSLQCIITENPGLIQLNAGVYKSNM-----LQGLQGL 423
 QY 329 -SQPSLSQSQWQPQ-----LQQLRDMGIQDELSLRPRCPVGTSGKOPWSSSLLEPHELP 383
 DB 424 FAGNPLOCEMQLPFTFLOQMN-----PDTLSMNSNPRAMQ 461
 QY 364 A-----SPEPPASCRNYCP 397
 DB 452 ALQIQQLIQCITENPGLIQLNAGVYKSNM-----LQGLQGL 466

RESULT 6
 US-10-293-000-2
 Sequence 2, Application US/10293000
 Publication No. US20030175278A1
 GENERAL INFORMATION:
 APPLICANT: Monteliro, Mervyn J.
 APPLICANT: Mah, Alex L.
 APPLICANT: Perry, George
 APPLICANT: Smith, Mark A.
 TITLE OF INVENTION: UBIQUILIN, A PRESENILIN INTERACTOR AND METHODS OF USING SAME
 FILE REFERENCE: 4115.175
 CURRENT APPLICATION NUMBER: US/10/293-000
 PRIOR APPLICATION NUMBER: 2003-04-01
 CURRENT FILING DATE: 2003-04-01
 PRIOR FILING DATE: 2001-11-13
 NUMBER OF SEQ ID NOS: 8
 SOFTWARE: Patent in version 3.2
 SEQ ID NO 2
 LENGTH: 595
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-293-000-2

Query Match 6.6%; Score 141.5; DB 12; Length 595;

Dd	237	PIPTSTTVQPSSSSAPTTSATSVQPSSSSSTPIPTTTSVQPSSSSAPITTSATSVPSPS	298
QY	224	FLUEGLSDDEDHFHNTRSTPSSSTPSSRPASLG---YSGAAGPRPITOSELATAALAS	280
Dd	297	SSSPZUSSTISVQPSSSSSSSTTTTSVQPSSSGSGAPTTSATSVQPSSSSPISSTISV	356
QY	281	TPESSSH:PTGTCT--QCHSSGTSRPMESGV-----OSGTPL-T	314
Dd	357	QPSSSSSSSTTTTSVQPSSSGSGAPTTSATSVQPSSSSVPTTSATSVRSSSSSSTPIPT	416
QY	315	NDLFSCA-QHALOASGPSSSQWQPQLQRMGIOCELSLRPCR-----PPVGT:SKO	369
Dd	417	TTSVQPSSSSVPTTSATSVQPSSSSTPIPTTTSVQPSSSSAPTTSATSVQPSSSSP	476
QY	370	PMSSSLLEPHELDASP-----EPPASCGRGYCPHEALMKRVPS	407
Dd	477	PISTTSVQPSSSSSSTTTTSVQPSSS--GSAPTTSATSVQPS	519
 RESULT 8 US-09-745-008-34			
Sequence 34, Application US/09745008			
Patent No. US2002037667A1			
GENERAL INFORMATION:			
APPLICANT: Chuenkova, Marina			
APPLICANT: Pereira, Mircio A.			
TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and			
TITLE OF INVENTION: Methods of Use Thereof			
FILE REFERENCE: 1324 1028-001			
CURRENT APPLICATION NUMBER: US/09745_008			
CURRENT FILING DATE: 2000-12-20			
PRIOR APPLICATION NUMBER: US 60/172,891			
PRIOR FILING DATE: 1999-12-20			
NUMBER OF SEQ ID NOS: 34			
SOFTWARE: FastSeq for Windows Version 4.0			
SEQ ID NO 34			
LENGTH: 1162			
TYPE: BNT			
ORGANISM: Trypanosoma cruzi			
US-09-745-008-34			
 Query Match 6.4%; Score 135.5; DB 10; Length 1162;			
Best Local Similarity 21.8%; Pred No. 0.1;			
Matches 98; Conservative 51; Mismatches 189; Indels 111; Gaps 18			
QY	6	WHLAWKADDP:TPKSILRLPETELGYSYGGYS:SLFKOLLAKLOESVDPDEL:D---	62
Dd	499	VHCVNFWAN -----KIGSVYDG-----ELLESGGTVVDGRTPDISH	537
QY	53	-----LIYGRKKAKDDTLDFYGIQP--GSTVHLRKSWPEPD	98
Dd	518	FVVGGRKSAEMPTISHVTNNVLLYNRQLNTEERT:FLSQDLICGTENHMDSDDSSAH	597
QY	ap	OKPE ECKKVAAKEFRVLHTALSSSSSYEAEPKMLSNKESLDIIVATPGUSSPIAL	157
Dd	598	STFTSTALUSSAHS:STPTDVSSAHSSTPTPADSSAHCPTPTPVDSHAHG:FSTPADSSAH	657
QY	158	GVCQKQGF:SVFAIFNM LDTLVPAHPALYNAIVLVHVSAGAMPDGTGTSRRSVP	214
Dd	658	GTPSTEV:ASAHSSTPTPVDSASHSTS-----TPVDSSAHGAPSTPADSSAHGTPSTP	711
QY	218	--SSVFVKGGEFLGCLSCDEDFHFNTRSTPSSSTPSSRPAS:GYGGAAGPRPITQSEL	272
Dd	712	VDSSAHG:P-----STPADSSAHSSTPTPADSSAHSSTFPADS-SAHSTFS-PVD-	761
QY	273	ATALALASTPSSSHPT:PGT----QQHSSGTFPMSSGVQS--GTPTNDLFSCALQHAI	326
Dd	762	SSAHCPTPTPADSSAHSSTPS:PADSSAHGCTPTPTVDSSAHSSTSTPTVDS-----	810
QY	327	QASGQ:SLQSQWQPQLQRMGIOCELSLRPCR-----PVGT:SKOPWSSSLLEPHEL	382
Dd	811	SAHTFTSTP-----VQSSAHSSTPTPTVDSSAHCPTPTPVDS-----AHST	851

```

Query Match      6.41; Score 135.5; DB 10; Length 1162;
Best Local Similarity 21.84; Pred. No. 0.1;
Matches 98; Conservative 51; Mismatches 189; Indels 111; Gaps 18;

QY      6 WHLAVK ADQPLTPKSLRLPELCEYSLGSGYS:SFLLKQLIAGLKQESVDPPELD:--- 62
Db      499 VHLVLTWAS :-----KISVVDG-----ELLESGQVTFDGTPIDISH 537
QY      4 :-----LIYCKRKAKDQDTLDFYGIQP--GSTVHYLRKSWPEPD 98
Db      518 FVYGVYKREKQMPITISHVTVNNVLLNPKQINTEEIRT:FLSQDLIQTENHWDSSSDSSAH 597
QY      99 OKPE KCKKVAAKREPRVLHTALHSSSSSYREAVFKVLSKESLDQIIVATPLGSSDPIAL 157
Db      598 STPSTPLAUSAHSSTPSTPVDSSSAHSTPSTPACSSAAGCTPSTPVDSSAAGCTPSTPADSSAH 657
QY      158 GVLCCKKQLF SVFAENM LELVPAHPA:LVNAIVLVCHSVAGSAPMPGCTDSSSRMPS-- 214
Db      659 CTPEFVLSAHSSTPSTPVDSSAHSSTP:-----TPVDSSAAGAPSTPADSSAAGCTPSTP 711
QY      215 --SSVPLMPGCLFEGC:SCDDEDFHNTBSTPSSSTPSRPAASLGYSGAAGCPRIIOSEL 272
Db      712 VDS:SAHGTP:-----STPADSSAHSSTPSTPADSSAHSSTPSTPADS:SAHSTPSTPVD- 761
QY      273 ATALALASTPESSSHTTPGCT:---QCHSGSGTSPWSSGVQS--GTPITNDLFSQALQHAL 326
Db      762 SSAHGCTPSTPADSSAHSSTPSTPADSSAAGCTPSTPVDSSAHSSTPSTPVDS:----- 810
QY      327 QASQF:SLGKQWQQLQGLADMGIQDELSLRPCRP:----PVGTSKQPMSSSLLERPEHEL 382
Db      811 SAHITPSTP:-----VQSSAHSSTPSTPVDSSAAGCTPSTPVDS:----AHST 851

```

QY 383 PASP-EPPASCRGYPWEALMKVPPSLPV 410
Db 852 PSTPADSSAHSSTPSTPADSSAHSHTPTSPV 590

RESULT 9
US-10-001-632A-2
; Sequence 2, Application US/1000-612A
; Publication No. US20020151492A1
; GENERAL INFORMATION:
; APPLICANT: Conklin, Carrell C.
; APPLICANT: Feinhaus, Andrew L.
; APPLICANT: Holderman, Susan D.
; TITLE OF INVENTION: Testis Specific Protein
; FILE REFERENCE: 99-1701
; CURRENT APPLICATION NUMBER: US/1000-612A
; PRIOR FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 60/128,210
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: 60/156,940
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: 69/541,9190
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-001-632A-2

Query Match 6.33; Score 134; DB 15; Length 655;
Best Local Similarity 19.43; Pred. No. 0.062;
Matches 86; Conservative 61; Mismatches 111; Indels 146; Gaps 16;

QY 39 SISFLKQLAGKLGESVPEPELIDLYCORKLKDDQTLDFYQ:QGSTVHVLRK----- 92
Db 42 TIOQLKEIS---QRFKAHPDQVLIFAGKILKDPDSLACQGVDRGLTVHVLKIQHRAH 98
QY 93 -----SWPEPQCKPEP-----VDKVAAMREFRVLHTAL--HSSSSY 126
Db 99 GNECPAASVPTQGPSFQSPGLQPSSYPADGPPAPFSLGLTGLSLGLAYRFPDQPSLM 158
QY 127 REAVFKM:SNKESLDQIIIVATPGLSSDPIALGVLCQKDLFS--VFADPNMDEL----- 178
Db 159 RQHV-----SVPEFVQLI-----DGFIPGLLSNTGLVRQLVLDNPHMQCLIQHNPEI 207
QY 179 --VPAHPALNNAIVLVHVSAGSAPKPGTDSRSSMPSSSYRDMPGGF-----LFEGLSD 231
Db 208 GHILNPEIWMOTLEFLNPNAMQEV---RSQDRVLSNLESIPGGYNVLCYTYTDM 263
QY 232 D-----ECCFHN---TRSTPSSSTPSSRPASLGYSAGAPRPITQSELATALASTPE 283
Db 264 PMLNAVQSGFGGNPATATTNATTTSCPSRX-----E 297
QY 284 SSSHTPTGTGHSSTGSPMSSGVSGTPTMTDLFS-----QA 321
Db 299 NCDPLNPNWSTH--GGSGRQSGCGGQDADF:SNRFPNFLGI:RLVYDYLQQLHENPQS 355
QY 322 LQHALQASGQPSLQSQWQPCQLRIMGICDELRLRCRPPVGTSGKOPWSSSLLEPHE 381
Db 356 LGTYLQGTASALSQSQ-----EPPPSVNRVPFSSPSQEPQS 392
QY 392 LPASPEPPASCRGYPWEALMKVP 405
Db 393 GQPLPEESVALKGRSSCPAFRLYP 416

RESULT 10
US-10-054-683-29
; Sequence 29, Application US/10054683
; Publication No. US2003004481A1
; GENERAL INFORMATION:

APPLICANT: Old, Lloyd J.
APPLICANT: Scanlan, Matthew J.
APPLICANT: Chen, Yao-Tseng
TITLE OF INVENTION: Cancer-Testis Antigens
FILE REFERENCE: LC461/7125 (JRV)
CURRENT APPLICATION NUMBER: US/10/054,683
CURRENT FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: US 60/280,718
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 60/285,154
PRIOR FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: US 60/327,432
PRIOR FILING DATE: 2001-10-05
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn version 3.1
SEQ ID NO 29
LENGTH: 655
TYPE: PRT
ORGANISM: Homo sapiens
US-10-054-683-29

Query Match 6.33; Score 134; DB 15; Length 655;
Best Local Similarity 19.43; Pred. No. 0.062;
Matches 86; Conservative 61; Mismatches 151; Indels 146; Gaps 16;

QY 39 SISFLKQLAGKLGESVPEPELIDLYCORKLKDDQTLDFYQ:QGSTVHVLRK----- 92
Db 42 TIOQLKEIS---QRFKAHPDQVLIFAGKILKDPDSLACQGVDRGLTVHVLKIQHRAH 98
QY 93 -----SWPEPQCKPEP-----VDKVAAMREFRVLHTAL--HSSSSY 126
Db 99 GNECPAASVPTQGPSFQSPGLQPSSYPADGPPAPFSLGLTGLSLGLAYRFPDQPSLM 158
QY 127 REAVFKM:SNKESLDQIIIVATPGLSSDPIALGVLCQKDLFS--VFADPNMDEL----- 178
Db 159 RQHV-----SVPEFVQLI-----DGFIPGLLSNTGLVRQLVLDNPHMQCLIQHNPEI 207
QY 179 --VPAHPALNNAIVLVHVSAGSAPKPGTDSRSSMPSSSYRDMPGGF-----LFEGLSD 231
Db 208 GHILNPEIWMOTLEFLNPNAMQEV---RSQDRVLSNLESIPGGYNVLCYTYTDM 263
QY 232 D-----ECCFHN---TRSTPSSSTPSSRPASLGYSAGAPRPITQSELATALASTPE 283
Db 264 PMLNAVQSGFGGNPATATTNATTTSCPSRX-----E 297
QY 284 SSSHTPTGTGHSSTGSPMSSGVSGTPTMTDLFS-----QA 321
Db 299 NCDPLNPNWSTH--GGSGRQSGCGGQDADF:SNRFPNFLGI:RLVYDYLQQLHENPQS 355
QY 322 LQHALQASGQPSLQSQWQPCQLRIMGICDELRLRCRPPVGTSGKOPWSSSLLEPHE 381
Db 356 LGTYLQGTASALSQSQ-----EPPPSVNRVPFSSPSQEPQS 392
QY 392 LPASPEPPASCRGYPWEALMKVP 405
Db 393 GQPLPEESVALKGRSSCPAFRLYP 416

RESULT 11
US-09-936-392-16
; Sequence 16, Application US/09836392
; Patent No. US20020173458A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Protein Tyrosine Kinase Receptor Polynucleotides, Polypeptides,
; FILE REFERENCE: PTO2001
; CURRENT APPLICATION NUMBER: US/09/836,392
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/US00/28066
; PRIOR FILING DATE: 2000-10-11
; PRIOR APPLICATION NUMBER: 60/159,542
; PRIOR FILING DATE: 1995-10-15

PRIOR APPLICATION NUMBER: 60/165,914
PRIOR FILING DATE: 1999-11-17
PRIOR APPLICATION NUMBER: 60/189,027
PRIOR FILING DATE: 2000-03-14
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 16
LENGTH: 1134
TYPE: PRT
ORGANISM: Homo sapiens
US-09-816-392-16

Query Match 6.0%; Score 128, DB 9; Length 841

Best Local Similarity 24.7%; Pred. No. 63;
Matches 69; Conservative 31; Mismatches 31; Indels 9; Gaps 14;

CY 195 SVAGSAPVORTS---SSASVSSSVYKMPGPHLEGLSDDECHPNTKSTPSSSTSS 251
DB 681 SLSSSSLSGSPHGLPARS-PHSVSTPDS-AVLCASSQS ---SSPASTTNS 730
CY 252 RPASLYSGAAGRPITOSLATALA-----ALASTESSHPTPTT 293
DB 731 -PAS---SASHH-RPSTHLSLSPK-HROVESAPKAGNIFRPLAHITSTQAFPP-L 785
CY 294 CGHSSST-----SPMSSTVQS HPTTHGLSQ 320
DB 786 PGHTVSSHTQSPKAKHSPPVVFEPKSAEPFAPLKSQSAEKUAGASAKAGA 845
CY 321 ALQHAIQASQPSLQSQWQ-----POCCAHMMIQIEHSL 357
DB 846 LRKHSLEV-GHPFRKOPFELALHSLARSUSHTPEVGLAPROVAVRQASQSPSLD 904
CY 358 RPCRPP-PVTSKQPNSSSLJHEHLEHLPASPIFASCRG 344
DB 905 -GADPLLPGASRPVSSKESKESQAGAEATTPKATPG 944

RESULT 12

US-09-816-860A-2

Sequence 2, Application US/99814460A

Patent No. US200001651A1

GENERAL INFORMATION:

APPLICANT: Meyers, Rachel

TITLE OF INVENTION: 26649, A No. US200001651A1

TITLE OF INVENTION: Uses of the

FILE REFERENCE: M1: 13

CURRENT APPLICATION NUMBER: US-09-816-860A-2

PRIOR APPLICATION NUMBER: 60/165,914

PRIOR FILING DATE: 1999-11-17

NUMBER OF SEQ ID NOS: 14

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 2

LENGTH: 881

TYPE: PRT

ORGANISM: Homo sapiens

US-09-816-860A-2

Query Match 6.0%; Score 127, DB 9; Length 841

Best Local Similarity 25.3%; Pred. No. 63;
Matches 77; Conservative 27; Mismatches 50; Indels 9; Gaps 14;

CY 158 GVLCQKDLFSVF---ADPNKEDTLVTAHPALQALNAIVLHLSAGHAPVWGTSSSRMPS 214
DB 512 GTLRKHHSFAFQPLPPTQGS-WP-----ATPEHSSSSAES 552
CY 215 SSYRDMPP-GGFLEGG-SCSDCHPNTKSTPSSSTSS 252
DB 553 SGCGTVPSSAGILFQGGSPG-DHSPKPKDVSAAVPAKPNNSQASLQAFCAAASH 611
CY 253 PASLGY-SGAGRPPTTCSLATALALASTESSHPTPTTPTGUG-HSSQTPHSGVQS 309
DB 612 QLSMSQPHNAAGSPHT---DRFAVKKPAPAPKPNPPPGIPGGQSSSGT-CHPPLSP 669

CY 310 GTPITNDFSOALQHALQASGQPSLQSQWQPMQLOLRDMGIQDBELSPRCPPVGT-SKQ 369
DB 669 KPTTRSP-SPPTQHTGPPGQPSAPSQLSAPRRY-----SSLSPICAPNHPPPQ 717
CY 370 PWSSSL-----LEEP-HELPAPEPPASCRGCPWALMKVP 405
DB 718 PPTCAITLMTKPNSSQPPNPMALPSEHGLEBQPSHTPTPTPST-----PPLGKQN 770
CY 406 PSLP 409
DB 771 PSLP 714

RESULT 13

US-10-146-473-50

Sequence 50, Application US/10146473

Publication No. US20010108888A1

GENERAL INFORMATION:

APPLICANT: Scania, Matthew

APPLICANT: Golt, Ivan

APPLICANT: Stockert, Elisabeth

APPLICANT: Gure, Ali

APPLICANT: Chen, Yao Tseng

APPLICANT: Old, Lloyd

TITLE OF INVENTION: Breast Cancer Antigens

FILE REFERENCE: US2001/0146473

CURRENT APPLICATION NUMBER: US/10/146,473

CURRENT FILING DATE: 2002-05-15

PRIOR APPLICATION NUMBER: US 60/291,150

PRIOR FILING DATE: 2001-05-15

NUMBER OF SEQ ID NOS: 82

SOFTWARE: Patent In Ver. 3.0

SEQ ID NO 50

LENGTH: 2971

TYPE: PRT

ORGANISM: Homo sapiens

US-10-146-473-50

Query Match 5.9%; Score 126; DB 15; Length 2971;

Best Local Similarity 22.7%; Pred. No. 211;

Matches 114; Conservative 49; Mismatches 178; Indels 162; Gaps 23;

CY 12 LABQPLTPKSL-----RLPETELGEYSLGGYISFLKQLIAGLQESVDP 58
DB 1068 LAPAPRPSGCLPAVINPRPTLTGRLPTTGTARAPMTTTLVRPL--KLHVS-RSP 1124
CY 59 ELIQLLYCGKHKDKDCTLQFYGIQ---GSTVHVLRKSMPEPDQKPEPVKVAAMREFRV 115
DB 1125 EV-----SASAPGAAPLTSSPLHV-PSSLPGPASSPMP:PNSSPL----- 1164
CY 116 LHTALSSSSSYREAVFKYLSNKESLDQI:VATPGSSDDPIALGVLDKDLFSVPADPNML 175
DB 1165 -----ASPVSTSVSP:SSSLPI:SVPTTLPAASAPLTIPISAPL----- 1204
CY 176 DTLVPAHPALVNAIVLHVSAGSAMPDGTSSSRMPSSSSYRDMPCGFLFGLDDEDD 235
DB 1205 -TVSASGPAULTSVTPPLAPVPAAPGPPSLQSPGASPSASALT-GLATAPSLSSQTP 1262
CY 236 FHP-----NTRSTPSSS---TPSSRPASLGYSGAGRPRTQS----- 270
DB 1263 GHPULLAP-SSHVPGLNSTVAPACSPVJVPASALAS-PFPSAPNPAPAOASLAPASSAS 1321
CY 271 -ELATLALASTPE-----SSSHPTPTGQGHSS-GT 300
DB 1322 QALATLAPMAAOTAILAPSPAPPAPLVLAPSPCAAPVLASSQTPVVPVAPSTPCT 1381
CY 301 SPMS-SVQSGTPTINDLFSQALQHALQASGQS-----LQSQMOPQL-----QQUR 346
DB 1382 SLASAPVAPPTVLPASSSTOTMLPAPVPSPLPSPASTOTLALAPALPTLQSSPSQTL 1441
CY 347 DMGLQDCE-----LSLRPCR-----PVCTSKOPWSSSLLEEHLEPA 384

```
Db 1442 SLGTGNGGPPPTOTLSLTPASSLVPTPAQT:SLAPGPPLGPTG---TLSLAPAPPAPPA 1498
Qy 385 SPBPASRCGVCPEALMKVPS 407
Db 1499 SPVGPA-----PAHTLTAFAS 1515

RESULT 14
US-10-029-386-32014
; Sequence 32014, Application US/10023386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hazzei, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBS USEFUL FOR Q
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/229,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32014
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008175.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.0
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: SWISSPROT HIT: P06640, EVALUE 2.00E-32
US-10-029-386-32014

Query Match 5.8%; Score 124.5; DB 12; Length 841;
Best Local Similarity 23.9%; Pred. No. 0.49;
Matches 77; Conservative 41; Mismatches 145; Indels 79; Gaps 14;

Qy 66 CGRLK22QQTLDYGIQPGSTVHLKRSWPEQCKPEPVKVAAYREFEVHTALHSSSS 125
Db 314 CWKQCADSGTLDSTNQ-----ELCTRS-----SSGTPVRFQTHVPMKSPSSGTFVSSG 364
Qy 126 YRE AVPKVLSNKSLEDQIVATFLESERIALGVLOQKHLFEVFAQPPVLDIVFAHPA 184
Db 365 FROSSQIKSVKNTQTSKTCQVETANQSR-----SAP-HEAHT HPA 476
Qy 185 LUNALVVLVHVAASAIWPTDSASQVWVKNLQVHNSVQVQVQVQVQVQVQVQVQV 244
Db 407 ---AQSSSRDSSSETFERSSTPSAT F-----PNNRIQT 439
Qy 245 SSSTPS-----SRPASLGVSGAAGRPITQSELATALALASTPSSHTPTTPTQCHSSGTPMSSG 296
Db 440 SSGTSPLEAPSSSTQLGAPHPCT-----KQISSNQPISRTQIFKAPSSBAQ 492
Qy 297 SS-GTSP-----MSGVQSGPTINDLFSQALHALQASGQFVQS--CWQPOLQQLREY 348
Db 493 SSPAPSSGTQLSSGAQAGSEIASSKTQTAAAT-QSSSFL--SSGVQSSPTCALAAI 551
Qy 349 GIQDELSLRPCRPVGTSGCP 370
Db 552 ELSTALSSDSRSSRSRTGRCP 573

RESULT 15
US-10-128-714-3176
; Sequence 3176, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
```

```
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: Methods of Use
; FILE REFERENCE: 10/82-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,990
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,162
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patent version 3.1
; SEQ ID NO 3176
; LENGTH: 1714
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3176

Query Match 5.8%; Score 124.5; DB 15; Length 1714;
Best Local Similarity 23.9%; Pred. No. 1.13;
Matches 68; Conservative 36; Mismatches 115; Indels 65; Gaps 13;

Qy 166 FSVFADPNMLDTVPAPALVNAIVLVLSVA-----GSAPMPGTDSSSRM----- 212
Db 1434 POLYDQKJLNTWSN-----NARFGVMGTIGAKDAIISDGAATQYDTCSPMQESAFIGT 1488
Qy 213 --PSSSY-----RPMQGFU-FEGLSDDEDDFHPTNSTPSSSTPSS---RPAISLY 258
Db 1489 PPESAFSP:ROAGAESPGCFTEYQTFGGFGGFSF-AATSPAGYSPSPSANTSPQY 1547
Qy 259 S-----GAAGPRPITQSELATALALASTPSSSHHTPTTQCHSSGTPMSSG 306
Db 1548 SPTSSVSP:SPMGITSPREYTSPGFSFA-SPSFAPTSFAYSPSPAYCOASPTSPSYSP 1606
Qy 307 VSGPTINDLFSQALHALQASGQFVQSQSQVQPOLQQLRDMGIQDELSLRPCRPVGT 366
Db 1607 TSPGESPTSNYSPTSPSPSPAS--PAF-SP7SP-----SYSP7SPAIGG 1648
Qy 367 SKQFASSSLESPHELPAPE-PPASCRGYCPEALMKVPSLP 409
Db 1649 AGRHLSPTSTSPKVTPTSPGWSPTSPCTYSTSPNFAGSP7SP 1692

Search completed: November 12, 2003, 06:44:24
Job time : 31 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: November 12, 2003, 06:36:13 / Search time 41 seconds
without alignments
403 429 Million cell updates/sec

Title: US-09-735-251-4

Perfect score: 2133

Sequence: 1 YSLSDWH:AVKLADQLTPY SFGYGVWALXVYRSLV 418

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 431858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents AA*

- 1: /cgn2_6/ptdata/1/1aa/5A_CWRB.pep*
- 2: /cgn2_6/ptdata/1/1aa/5B_CWRB.pep*
- 3: /cgn2_6/ptdata/1/1aa/6A_CWRB.pep*
- 4: /cgn2_6/ptdata/1/1aa/6B_CWRB.pep*
- 5: /cgn2_6/ptdata/1/1aa/PTB5_CWRB.pep*
- 6: /cgn2_6/ptdata/1/1aa/backfiles.pep*

Pred. No. is the number of residues predicted by domain 1. Have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2133	100.0	418	US-09-735-251-4	Sequence 1, Appl
2	2129	99.8	418	US-09-735-251-4	Sequence 2, Appl
3	139.5	6.3	1393	US-09-735-251-4	Sequence 3, Appl
4	139.5	6.3	1393	US-09-735-251-4	Sequence 4, Appl
5	126	5.9	126	US-09-735-251-4	Sequence 5, Appl
6	126	5.9	126	US-09-735-251-4	Sequence 6, Appl
7	126	5.9	126	US-09-735-251-4	Sequence 7, Appl
8	122.5	5.7	816	US-09-735-251-4	Sequence 8, Appl
9	122.5	5.7	1187	US-09-735-251-4	Sequence 9, Appl
10	122.5	5.7	1187	US-09-735-251-4	Sequence 10, Appl
11	122.5	5.7	1187	US-09-735-251-4	Sequence 11, Appl
12	122.5	5.7	1210	US-09-735-251-4	Sequence 12, Appl
13	122.5	5.7	1210	US-09-735-251-4	Sequence 13, Appl
14	122.5	5.7	1210	US-09-735-251-4	Sequence 14, Appl
15	121	5.7	1060	US-09-735-251-4	Sequence 15, Appl
16	118	5.5	345	US-09-735-251-4	Sequence 16, Appl
17	118	5.5	345	US-09-735-251-4	Sequence 17, Appl
18	118	5.5	878	US-09-735-251-4	Sequence 18, Appl
19	117	5.5	543	US-09-735-251-4	Sequence 19, Appl
20	117	5.5	577	US-09-735-251-4	Sequence 20, Appl
21	117	5.5	1129	US-09-735-251-4	Sequence 21, Appl
22	117	5.5	1646	US-09-735-251-4	Sequence 22, Appl
23	117	5.5	1647	US-09-735-251-4	Sequence 23, Appl
24	117	5.5	1649	US-09-735-251-4	Sequence 24, Appl
25	117	5.5	1650	US-09-735-251-4	Sequence 25, Appl
26	117	5.5	1678	US-09-735-251-4	Sequence 26, Appl
27	117	5.5	1679	US-09-735-251-4	Sequence 27, Appl

28	117	5.5	1681	4	US-09-535-008-77	Sequence 77, Appl
29	117	5.5	1682	4	US-09-535-008-77	Sequence 73, Appl
30	116.5	5.5	706	2	US-08-074-967-2	Sequence 2, Appl
31	116.5	5.5	706	2	US-08-553-541B-2	Sequence 2, Appl
32	116.5	5.5	706	3	US-09-268-202-2	Sequence 2, Appl
33	116.5	5.5	706	5	PCT-US94-06669-2	Sequence 2, Appl
34	115.5	5.4	2414	1	US-08-227-546-2	Sequence 2, Appl
35	115.5	5.4	2414	5	PCT-US95-04682-2	Sequence 2, Appl
36	115	5.4	907	3	US-08-783-774-2	Sequence 2, Appl
37	115	5.4	907	4	US-09-328-599A-1	Sequence 1, Appl
38	115	5.4	907	5	PCT-US95-04611A-19	Sequence 19, Appl
39	114.5	5.4	1142	2	US-08-993-118-7	Sequence 7, Appl
40	114.5	5.4	1142	3	US-08-845-528C-7	Sequence 7, Appl
41	114.5	5.4	1142	4	US-09-066-281B-7	Sequence 7, Appl
42	114	5.3	1402	4	US-09-125-635-12	Sequence 12, Appl
43	113.5	5.3	2442	4	US-09-514-247A-10	Sequence 10, Appl
44	113	5.3	749	4	US-08-997-685A-10	Sequence 10, Appl
45	113	5.3	1321	2	US-08-317-310A-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1
US-08-725-758A-4
Sequence 4, Application: US/09725758A
Patent No. 6160109
GENERAL INFORMATION:
APPLICANT: Reed, Guy
APPLICANT: Client, Christophe Y.
TITLE OF INVENTION: NOVEL PLATELET ACTIVATION PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110 2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,758A
FILING DATE: 04-OCT-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 60/005,074
FILING DATE: 4 OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/C20001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-6906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-725-758A-4

Query Match: 100.0% Score 2133; DB 3; Length 410;
Best Local Similarity: 100.0%; Pred. No. 4.6e-181;
Matches 410; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

OY : MSLS-WH-AVK-ADQP-PTKSLRLPETELGYSISGLKOLIAKGLQESVDPDEL 60
|||||

```

Db 1 MSLSCHHLAVKADQPLTPKSIIRLPETELGEYSLSGGYSISFKQIAGKLCESVDPDEL 60
Qy 61 IDLIYCGRLKDDQTLDFYGIQFGSTVHVLRKSWPEPDKPEPDKVAAAREFRVLTAL 120
Db 61 IDLIYCGRLKDDQTLDFYGIQFGSTVHVLRKSWPEPDKPEPDKVAAAREFRVLTAL 120
Qy 121 HSSSYREAVENKLSKESLDQ::VATPGUSSDP::ALGVYQDKLFSVADPNMLDTLVP 180
Db 121 HSSSYREAVENKLSKESLDQ::VATPGUSSDP::ALGVYQDKLFSVADPNMLDTLVP 180
Qy 181 AHPALVNAIVLVHVSAGSAPMPCTOSSSRSMSSSSSYRDMPGGFLFEGLSDDDDHPNT 240
Db 181 AHPALVNAIVLVHVSAGSAPMPCTOSSSRSMSSSSSYRDMPGGFLFEGLSDDDDHPNT 240
Qy 241 RSTPSSSTRSRPASLGYSAGAGPRP::TQSELATALALASTPSSSHTPTGTQGHSSGT 300
Db 241 RSTPSSSTRSRPASLGYSAGAGPRP::TQSELATALALASTPSSSHTPTGTQGHSSGT 300
Qy 301 SPMSGVQSGTPTNDLFSQALQHALQASQPSLOSQWOPQLOQLDMGIQDDLSLRPC 360
Db 301 SPMSGVQSGTPTNDLFSQALQHALQASQPSLOSQWOPQLOQLDMGIQDDLSLRPC 360
Qy 361 RPPVGTSGKQPMWSSSLLEEPHELSPASPEPPASCRGYCPWEALMKVPPSLP 409
Db 361 RPPVGTSGKQPMWSSSLLEEPHELSPASPEPPASCRGYCPWEALMKVPPSLP 409

RESULT 2
US-08-725-758A-2
; Sequence 2, Application US/0674575A
; Patent No. 6160108
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy
; TITLE OF INVENTION: NOVEL FLAT-LET ACTIVATION PHARMIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson LLP
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110 2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTA, Seq, Gaps, Gaps
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/04-11-758A
; FILING DATE: 04-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 62/005,014
; FILING DATE: 06-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,813
; REFERENCE/DOCKET NUMBER: 25431/229A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5170
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 426 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-725-758A-2

```

Query Match 99.84, Score 2129, US 3, Length 426,
Best Local Similarity 100.00, Freq. No. 1 to 180.

```

Matches 409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MSLSCHHLAVKADQPLTPKSIIRLPETELGEYSLSGGYSISFKQIAGKLCESVDPDEL 60
Db 1 MSLSCHHLAVKADQPLTPKSIIRLPETELGEYSLSGGYSISFKQIAGKLCESVDPDEL 60
Qy 61 IDLIYCGRLKDDQTLDFYGIQFGSTVHVLRKSWPEPDKPEPDKVAAAREFRVLTAL 120
Db 61 IDLIYCGRLKDDQTLDFYGIQFGSTVHVLRKSWPEPDKPEPDKVAAAREFRVLTAL 120
Qy 121 HSSSYREAVENKLSKESLDQ::VATPGUSSDP::ALGVYQDKLFSVADPNMLDTLVP 180
Db 121 HSSSYREAVENKLSKESLDQ::VATPGUSSDP::ALGVYQDKLFSVADPNMLDTLVP 180
Qy 181 AHPALVNAIVLVHVSAGSAPMPCTOSSSRSMSSSSSYRDMPGGFLFEGLSDDDDHPNT 240
Db 181 AHPALVNAIVLVHVSAGSAPMPCTOSSSRSMSSSSSYRDMPGGFLFEGLSDDDDHPNT 240
Qy 241 RSTPSSSTRSRPASLGYSAGAGPRP::TQSELATALALASTPSSSHTPTGTQGHSSGT 300
Db 241 RSTPSSSTRSRPASLGYSAGAGPRP::TQSELATALALASTPSSSHTPTGTQGHSSGT 300
Qy 301 SPMSGVQSGTPTNDLFSQALQHALQASQPSLOSQWOPQLOQLDMGIQDDLSLRPC 360
Db 301 SPMSGVQSGTPTNDLFSQALQHALQASQPSLOSQWOPQLOQLDMGIQDDLSLRPC 360
Qy 361 RPPVGTSGKQPMWSSSLLEEPHELSPASPEPPASCRGYCPWEALMKVPPSLP 409
Db 361 RPPVGTSGKQPMWSSSLLEEPHELSPASPEPPASCRGYCPWEALMKVPPSLP 409

RESULT 3
US-08-545-860D-55
; Sequence 55, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute leukemia
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 60401401s
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D
; FILING DATE: 07 MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/662,443

```

FILING DATE: 14-MAY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/971,094
 FILING DATE: 30-OCT-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/886,839
 FILING DATE: 27-MAY-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/805,091
 FILING DATE: 11-DEC-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: DeLuca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: TCU 1262
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 55:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1093 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ANTI-SENSE: NO

US-08-545-860D-55

Query Match 6.5%; Score 139.5; DB 3; Length 1093;
 Best Local Similarity 21.4%; Pred. No. 0.0014;
 Matches 108; Conservative 48; Mismatches 181; Indels 167; Gaps 20;

QY	11	KIADQPLT---	PKSLRLPETELGYSYSGGYS:SELKOLIAG	-----	KJESVP 56
DB	590	RLSRSPFTSLPSSAS:STTQV--	FSLAGSTFSLPSTHIFGTMGAVNPALSSAESHT	647	
QY	57	DEELIDLVYCGKDKDDTDFYV	CGQSTVHV-----	-----	LKSWPEPDQKE 102
DB	648	EPDLEDCFRGRTGTSQESLS--	SVSPISLFDQTSAPCCGGQDPAAGTTNWEQ	705	
QY	103	PVDK-----	VAAMREFVLHTALHSSSYSEAVFKMLSNKESLDQIIIVATPGLSSDPI	155	
DB	706	LLEKQDGEAGNIVVEY-KALHA	CKENQRLCEQILSLTAKKERLQILNV-----	755	
QY	156	ALGVLCQDLFSVADPNVLDL	VPALVNAIVVLHVSAGSAPMPCD-----	SSSRM 212	
DB	756	-----	-----	-----	-----
QY	213	PSSSYRMYEGGFLEGLSDDED	DFHNTHTSPSSSTPPSRPASLGYSAGGPRPTOSSEL	272	
DB	798	PKSSSG-----	LNSLSTSSD-----	PHSCPPSRSSSLSFHTPTPLPQQSP	844
QY	273	AT-ALALASTPSSSHTP-----	TPGTQGHSSGTSYPMSSGVQSGTPTINDLFSQALQHA	325	
DB	845	ATPLALPGAPAPLPQPGNGLGRAP	GAAG--LGAMPNABGLGLAGSGGLPGLNGLGG	902	
QY	326	-----	QASQPSLQ-----	-----	SOMQPOLQQLRDMQIDDEL 357
DB	903	LNGAAAFNPASLSQAGGAPTLOL	PGCNLSLTQORHLIQOOQOQLOO-----	QLLA 955	

RESULT 4

PCT-US94-04496-55

Sequence 55, Application PC/TUS9404496

GENERAL INFORMATION:

APPLICANT: Croce, Carlo
 APPLICANT: Canaan, Eli
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1
 NUMBER OF SEQUENCES: 86
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
 ADDRESSEE: Norris
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0. Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/04496
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: DeLuca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: TCU-1242
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 55:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1093 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ANTI-SENSE: NO

PCT-US94-04496-55

Query Match 6.5%; Score 139.5; DB 5; Length 1093;
 Best Local Similarity 21.4%; Pred. No. 0.0014;
 Matches 108; Conservative 48; Mismatches 181; Indels 167; Gaps 20;

QY	11	KIADQPLT---	PKSLRLPETELGYSYSGGYS:SELKOLIAG	-----	KJESVP 56
DB	590	RLSRSPFTSLPSSAS:STTQV--	FSLAGSTFSLPSTHIFGTMGAVNPALSSAESHT	647	
QY	57	DEELIDLVYCGKDKDDTDFYV	CGQSTVHV-----	-----	LKSWPEPDQKE 102
DB	648	EPDLEDCFRGRTGTSQESLS--	SVSPISLFDQTSAPCCGGQDPAAGTTNWEQ	705	
QY	103	PVDK-----	VAAMREFVLHTALHSSSYSEAVFKMLSNKESLDQIIIVATPGLSSDPI	155	
DB	706	LLEKQDGEAGNIVVEY-KALHA	CKENQRLCEQILSLTAKKERLQILNV-----	755	
QY	156	ALGVLCQDLFSVADPNVLDL	VPALVNAIVVLHVSAGSAPMPCD-----	SSSRM 212	
DB	756	-----	-----	-----	-----
QY	213	PSSSYRMYEGGFLEGLSDDED	DFHNTHTSPSSSTPPSRPASLGYSAGGPRPTOSSEL	272	
DB	798	PKSSSG-----	LNSLSTSSD-----	PHSCPPSRSSSLSFHTPTPLPQQSP	844
QY	273	AT-ALALASTPSSSHTP-----	TPGTQGHSSGTSYPMSSGVQSGTPTINDLFSQALQHA	325	
DB	845	ATPLALPGAPAPLPQPGNGLGRAP	GAAG--LGAMPNABGLGLAGSGGLPGLNGLGG	902	
QY	326	-----	QASQPSLQ-----	-----	SOMQPOLQQLRDMQIDDEL 357
DB	903	LNGAAAFNPASLSQAGGAPTLOL	PGCNLSLTQORHLIQOOQOQLOO-----	QLLA 955	

Query Match 5.7% Score 122.5; DB 4; Length 816;
Best Local Similarity 22.7%; Pred No. 0.11;
Matches 114; Conservative 49; Mismatches 178; Indels 12; Gaps 23;
DB 12 LACDPLTKSIL-----RLFETELREYSLSGYSIFSLKGLIAGKLSGVPDP 58
1214 LAPARPSSGLPAVLNPPPTTWRURLETPFLICARAPKPTTLVRPLL KLVS-PS 1270
QY 59 ELIDLIYGRKLKQDQTLFYQI- GSTVFLRKSWPDPKFEVDVNAAMREFV 115
DB 1271 EV-----SASAPCAAPLTSSILNV PSEDEGPASSEMPLESSPL 1310
QY 116 LHTALHSSSYREAVFKLVKSPESLQIIVATILSSDPLALGVLCORLSSVADNKL 175
DB 1311 -----ASPVSTSVVLSSELP-SVPTTLTAPASAPLTIPASAP 1350
QY 176 DTLVPAHPALNAIVLVHSAVAAPVGTSSSSSYSPSSYKMPJ3FLPEGLSDREDD 235
DB 1351 -TVSASGPAALLTSVTPPIAPVVFAPAPGPPSOPKASASALUL-GLATAPS-SSQIF 1408
QY 236 FHP-----NTRSFSSS TPSPASASGYSAGAPRFTQS----- 270
DB 1409 GHPLLAPTSSHPVGLNATVAFACNVLVPASALAS-PPPSAENPAPACASLLAPASAL 1467
QY 271 -ELATALASTPE-----SSSTHTTPTQCHS GT 300
DB 1469 QALATPLAPMAAPQTA--APAPPLAPLPVATSPGAAPVLAASQTPVPAVSS-PTG 1527
QY 301 SPMS-SGVOSGTPIIND-FSCALQALCASQPS-----LCSQKPG-----QIL 346
DB 1528 SLASASPVATPVLAPSTQMLFAPVPSLAPSPASTOTLALAPALPTGSSSPCTL 1587
QY 347 DMGICQDE-----LSHPPH-----FPVTSKQPASSSLEDEPHLEDA 164
DB 1588 SLGTGNPGCPPTQTLSTPSSASVETPAQTLNAPGPPGHTG TLSAPAPPLAPA 1644
QY 385 SPERPASRGVCPNEMALVKVPS 407
DB 1645 SPVGF- ----PAHTLTAPAS 1741

RESULT 8
US-09-735-251-4-12
Sequence 12, Application US/09266225D
Patent No. 6573364
GENERAL INFORMATION:
APPLICANT: Nandabalan, Krishna
APPLICANT: Kishore, Stephen
APPLICANT: TCGenev, Veeraz
TITLE OF INVENTION: Isolation and characterization of human and monkey lymph
TITLE OF INVENTION: Syndrome (HIS) Protein Defects and HES Protein
TITLE OF INVENTION: Interacting Proteins
FILE REFERENCE: 15966-523
CURRENT APPLICATION NUMBER: US/09266225D
CURRENT FILING DATE: 1999-03-10
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 12
LENGTH: 816
TYPE: PRT
ORGANISM: Homo sapiens
US-09-266-225D-12
Query Match 5.7% Score 122.5; DB 4; Length 816;
Best Local Similarity 25.1%; Pred No. 0.01;
Matches 85; Conservative 39; Mismatches 116; Indels 99; Gaps 18;
QY 90 LRKSWPEQKPEPVKVA- ---MREFRVLHTA-HSSSSSYREAVFKLSKESLQIIVA 146
DB 491 LRDPGAPLEAPKPKVPTACQREEREKPRQERAKERE---KRPQERERKRCAGA 547
QY 147 TPGUSSDPIALGVLDKDLFSVFADPNMLDTLV-----PAHPALVATLVLVHSAVAGAPM 202

DB 548 SGGPTDPIAGLVLSN-----DRSLERWTRMARPAAPALT-----SVPAPAPA 592
QY 203 PGTDSSSRKSVSSYRDMPCGFLPEGLSDDEDDFHNTTRSTPSSSTPSSRPA-----SL 256
DB 593 P-TPTTPVQPTSP-----PQPVAPQPG-----PQPSAGSTSGVPQPCPPGPAP 640
QY 257 GYSGAAGPRI-TQSELATALALASTPSSSHHTPTTGTQCHS-----SG 299
DB 641 HPTGPEGPIVFPAPPQTATSTSLA---AOSLVPPGCLPGSSTPGVLPVFPGLPPPDAG 697
QY 300 TSMSSSOVQSGTP-----ITNDLFSQALQHALQA--SCOP-----SLOSOKPQLOOLR 346
DB 698 GACQSS--XSESQVNVLTQOLSKSQVEDPLPVFSGTPKSGAGYGVGFDLFEFLNQSF 755
QY 347 DMGICQD-----DELSLRCPRPVGTSGKOPWSSSLJEE 378
DB 756 DMGVADGPPGQADASL-----SASLLAD 780

RESULT 9
US-08-320-559-28
Sequence 28, Application US/08320559
Patent No. 5633135
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
TITLE OF INVENTION: All-1 Region
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/320,559
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/062,443
FILING DATE: 14 MAY 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/971,094
FILING DATE: 30-OCT-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/888,830
FILING DATE: 27-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/805,093
FILING DATE: 11-DEC-91
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-0855
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-320-559-28

Query Match 5.7%; Score 122.5; DB 1; Length 1187;
 Best Local Similarity 21.1%; Pred. No. 0.053;
 Matches 88; Conservative 52; Mismatches 171; Indels 107; Gaps 18;

QY 8 LAVKLCAPLTKPSIL-----RPFTELGEYSIGGYSIFLKQLIA-----GKQESYPD 57
 DB 226 LAVKVKKE-TPQSLVAPAQPSQTFPPSPKSVAMQCKPTAYVRPMQCCQAPSES 284
 QY 58 PELADLIYGRKKDD---QTLDFYGIQPGSTVHLRKSWPEPCQPEPVKVKVAMREF- 113
 DB 285 PEL-----KPLPEDYRQOTTEKTLKVPAAKUTLKLKMPQSVQEQYSNEVHCVEIL 337
 QY 114 -RVLH-----TALHSSSYREAVFQMLSNKESLQCIIVATGLSSDPIALGVLCQDLF 166
 DB 338 KEMTHSNPPPTTAHTTESTAEPSKEP FPKTS-----NSQGTSSMLEDLQ----- 405
 QY 167 SVFADPNMLDTLVFAHPALYNAVLVHSHVAGSAFVPGTDSRSSRMPSSSYDWPGGELF 226
 DB 374 SVTONKQYDTSSKTHS-----NSQGTSSMLEDLQ----- 405
 QY 227 EGLSDDED-DHPNTRTPSSSTPSSPPASLGVSGAAGPRPTQSELATALA-LASTPES 284
 DB 406 --LSDSDSDSEQTPKEKPPSSAPPAPQSL-----PEPVASHSSASESTSDSDS 456
 QY 285 SSHTPTPTQCHSSGTSPMSSGVQSGTPTINDLF--SQALQHALQAGOPSLQSQWQPOL 342
 DB 457 SSDSESSSSSDSENEPELETAPAEPEPTTNKWLQCNLTKVQSPAAPPSPRSTPEPR 516
 QY 343 QCLRDYVHDDDELRLPCR----PPVGTSKOPKSSSLLEPH-----ELPASPEPP 389
 DB 517 RHPESKSSDSATSCHESKSKDPSPSSSKAPRAPP--EAPHFGKRSCKSPAQCEPP 572

RESULT 13

US-08-545-8600-28
 Sequence 28, Application US/08545860
 Patent No. 6043140

GENERAL INFORMATION:

APPLICANT: Ciba, Cario
 APPLICANT: Ciba, Cario
 TITLE OF INVENTION: Diagnostic, Therapeutics and Methods
 TITLE OF INVENTION: Diagnostic, Therapeutics and Methods
 TITLE OF INVENTION: Diagnostic, Therapeutics and Methods
 TITLE OF INVENTION: Diagnostic, Therapeutics and Methods
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:

ADDRESSER: Woodcock, Washington, D.C., 20540-0101
 STREET: One Liberty Place, 40th floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS DOS
 SOFTWARE: Patent Release #100, Version # 1.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08545860
 FILING DATE: 07-MAR-1996
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94-04496-28
 FILING DATE: 22-APR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94-04496-28
 FILING DATE: 09-DEC-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/320,559
 FILING DATE: 19-OCT-1994
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/320,559
 FILING DATE: 11-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/062,443
 FILING DATE: 14-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/971,394
 FILING DATE: 10-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/888,839
 FILING DATE: 27-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/805,092
 FILING DATE: 11-DEC-1991

ATTORNEY/AGENT INFORMATION:

NAME: Deluca, Esq., Mark
 REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER:

REFERENCE/DOCKET NUMBER: 130-1262
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO:

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 1187 amino acids
 TYPE: amino acid
 TOPOLOGY: linear

MOLECULE TYPE:

MOLECULE TYPE: protein

US-08-545-8600-28

Query Match

Best Local Similarity 21.1%; Pred. No. 0.053;
 Matches 88; Conservative 52; Mismatches 171; Indels 107; Gaps 18;

QY 8 LAVKLCAPLTKPSIL-----RPFTELGEYSIGGYSIFLKQLIA-----GKQESYPD 57
 DB 226 LAVKVKKE-TPQSLVAPAQPSQTFPPSPKSVAMQCKPTAYVRPMQCCQAPSES 284
 QY 58 PELADLIYGRKKDD---QTLDFYGIQPGSTVHLRKSWPEPCQPEPVKVKVAMREF- 113
 DB 285 PEL-----KPLPEDYRQOTTEKTLKVPAAKUTLKLKMPQSVQEQYSNEVHCVEIL 337
 QY 114 -RVLH-----TALHSSSYREAVFQMLSNKESLQCIIVATGLSSDPIALGVLCQDLF 166
 DB 338 KEMTHSNPPPTTAHTTESTAEPSKEP FPKTS-----NSQGTSSMLEDLQ----- 405
 QY 167 SVFADPNMLDTLVFAHPALYNAVLVHSHVAGSAFVPGTDSRSSRMPSSSYDWPGGELF 226
 DB 374 SVTONKQYDTSSKTHS-----NSQGTSSMLEDLQ----- 405
 QY 227 EGLSDDED-DHPNTRTPSSSTPSSPPASLGVSGAAGPRPTQSELATALA-LASTPES 284
 DB 406 --LSDSDSDSEQTPKEKPPSSAPPAPQSL-----PEPVASHSSASESTSDSDS 456
 QY 285 SSHTPTPTQCHSSGTSPMSSGVQSGTPTINDLF--SQALQHALQAGOPSLQSQWQPOL 342
 DB 457 SSDSESSSSSDSENEPELETAPAEPEPTTNKWLQCNLTKVQSPAAPPSPRSTPEPR 516
 QY 343 QCLRDYVHDDDELRLPCR----PPVGTSKOPKSSSLLEPH-----ELPASPEPP 389
 DB 517 RHPESKSSDSATSCHESKSKDPSPSSSKAPRAPP--EAPHFGKRSCKSPAQCEPP 572

RESULT 11

PCT-US94-04496-28
 Sequence 28, Application PCT/US9404496
 GENERAL INFORMATION:

APPLICANT:

APPLICANT: Ciba, Cario
 APPLICANT: Ciba, Cario

TITLE OF INVENTION:

TITLE OF INVENTION: Diagnostic, Therapeutics and Methods
 TITLE OF INVENTION: Diagnostic, Therapeutics and Methods

NUMBER OF SEQUENCES:

NUMBER OF SEQUENCES: 86
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
 ADDRESSEE: Norris
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS DOS
 SOFTWARE: Patent Release #100, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/74496
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: DeLuca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: RU 1042
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3300
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 28
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1187 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT US94-04496-28

Query Match 5.78; Score 122.5; PB 1; Length 1187;
 Best Local Similarity 21.1%; Pred. No. 0.054;
 Matches 88; Conservative 52; Mismatches 171; Indels 107; Gaps 18;
 QY 8 LAVKLADQLTPKSIJ-----RLPETELGEYSIGGVYSISFLKOLIA-----GKLOESVDP 57
 DB 226 LAVKVDKE TPQDSLAVAPQPSOTFPFPLPSKSVAMQOKPTAYVRPMDCCQAPSES 307
 QY 58 PELIDLLYCGRKLDK---QTLDFYGIQPGSTVHLVRSWPEPDOKPEPVDKVAAMREF- 113
 DB 308 PEL-----KELPDYRQQTFEKDLKVPKAKLTKLKPVSQSVQYTSNEVHCVEIL 360
 QY 114 -RVLH-----TALHSSSYREAVFKM--SNKESLDQIIIVATPGLSSDPALGLVDKDLF 166
 DB 361 KEMTHSWPPELTATHTPSTAEPSKPE-FPTKDS-----QHVS 396
 QY 167 SVFADPNV-DLVPAPALVNAVLVLSHVSAGSAMPQGTDSRSRMPSSSYRDMVGGFLP 226
 DB 397 SVTONKQVDTSSKTHS-----NSQQTSSMLLEDLQ----- 428
 QY 227 EGLSDDED-DHPNTRSTPSSSRPASLGYSCAAGRPRTQSELATALA-LASTPES 284
 DB 429 ---SSSESSCSSEQTPEKPPSSAPSAFQSG-----PEPVASHSSASESTSDS 479
 QY 285 SSGHTPTGTHSSGTSPMSGVQSGTPTITNDF--SQALHALQASQPSLQSQWQPOL 342
 DB 480 SSESSESSSSSENEPLETAPPEPEPT--NKWOLDNWLTKVSPQAPPEGPRSTPEPR 539

PRESET 12
 US-09-735-251-4-26
 Sequence 26, Application US/08320559
 Patent No. 5633135
 GENERAL INFORMATION:
 APPLICANT: Croce, Carlo
 APPLICANT: Caranani, Eli
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
 TITLE OF INVENTION: Detection and Treatment of Acute Dehydration

TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
 TITLE OF INVENTION: All-1 Region
 NUMBER OF SEQUENCES: 44
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135r1s
 STREET: One Liberty Place - 46th Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: PC-DOS
 SOFTWARE: MICROPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/320,559
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/062,443
 FILING DATE: 14 MAY 1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/971,094
 FILING DATE: 30-OCT-92
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/888,830
 FILING DATE: 27-MAY-92
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/805,093
 FILING DATE: 11-DEC-91
 ATTORNEY/AGENT INFORMATION:
 NAME: DeLuca, Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: 707-0855
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3300
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 26
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1210 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-320-559-26

Query Match 5.78; Score 122.5; PB 1; Length 1210;
 Best Local Similarity 21.1%; Pred. No. 0.054;
 Matches 88; Conservative 52; Mismatches 171; Indels 107; Gaps 18;
 QY 8 LAVKLADQLTPKSIJ-----RLPETELGEYSIGGVYSISFLKOLIA-----GKLOESVDP 57
 DB 249 LAVKVDKE TPQDSLAVAPQPSOTFPFPLPSKSVAMQOKPTAYVRPMDCCQAPSES 307
 QY 58 PELIDLLYCGRKLDK---QTLDFYGIQPGSTVHLVRSWPEPDOKPEPVDKVAAMREF- 113
 DB 308 PEL-----KELPDYRQQTFEKDLKVPKAKLTKLKPVSQSVQYTSNEVHCVEIL 360
 QY 114 -RVLH-----TALHSSSYREAVFKM--SNKESLDQIIIVATPGLSSDPALGLVDKDLF 166
 DB 361 KEMTHSWPPELTATHTPSTAEPSKPE-FPTKDS-----QHVS 396
 QY 167 SVFADPNV-DLVPAPALVNAVLVLSHVSAGSAMPQGTDSRSRMPSSSYRDMVGGFLP 226
 DB 397 SVTONKQVDTSSKTHS-----NSQQTSSMLLEDLQ----- 428
 QY 227 EGLSDDED-DHPNTRSTPSSSRPASLGYSCAAGRPRTQSELATALA-LASTPES 284
 DB 429 ---SSSESSCSSEQTPEKPPSSAPSAFQSG-----PEPVASHSSASESTSDS 479
 QY 285 SSGHTPTGTHSSGTSPMSGVQSGTPTITNDF--SQALHALQASQPSLQSQWQPOL 342
 DB 480 SSESSESSSSSENEPLETAPPEPEPT--NKWOLDNWLTKVSPQAPPEGPRSTPEPR 539

343 QQLRDMG:QDDELSPR-----PVTGTSKQWSSLSLEERH-----ELPSPPEP 349
 540 RHPEKSGSDSATSQESESKEPPPSKKAFAFAP--EAPHGKRSQVFAQOEPP 595
 RESULT 13
 US-08-545-8600-26
 ? Sequence 26, Application US/0845458600
 ? Patent No. 6042140
 ? GENERAL INFORMATION:
 ? APPLICANT: Crocco, Carlo
 ? APPLICANT: Canabadi, Eli
 ? TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
 ? TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
 ? TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1 Region
 ? NUMBER OF SEQUENCES: 94
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
 ? ADDRESS: No. 604214015
 ? STREET: One Liberty Place, 46th floor
 ? City: Philadelphia
 ? STATE: Pennsylvania
 ? COUNTRY: USA
 ? ZIP: 19103
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? OPERATING SYSTEM: IBM PC compatible
 ? SOFTWARE: Patent in Release #1.0, Version #1.25
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/0845458600
 ? FILING DATE: 07-MAR-1994
 ? CLASSIFICATION: 435
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: PCT/US94/04496
 ? FILING DATE: 22-APP-1994
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: PCT/US94/04496
 ? FILING DATE: 09-DEC-1992
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: US 07/971,094
 ? FILING DATE: 19 OCT-1994
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: US 07/989,849
 ? FILING DATE: 27-MAY-1992
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: US 07/989,849
 ? FILING DATE: 11-DEC-1992
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: DeLuca Esq., Mark
 ? REGISTRATION NUMBER: 33,429
 ? REFERENCE/DOCKET NUMBER: TJU-1262
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (215) 568-3100
 ? TELEFAX: (215) 568-3439
 ? INFORMATION FOR SEQ ID NO: 26:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 1210 amino acids
 ? TYPE: amino acid
 ? TOPOLOGY: linear
 ? N-TERMINUS TYPE: Protein
 ? US-08-545-8600-26

Query Match 5.71; Score 122.5; DB 3; Length 1210;
 Best Local Similarity 21.11; Pred. No. C.054;
 Matches 89; Conservative 52; Mismatches 171; Indels 107; Gaps 18;
 QY 8 LAVKLADQELTPKSL-----RUPETELGEYSUGYGISFLKOLIA-----GKQESVVD 57
 DB 245 LAVKVKCKE TPQDSLVAQAQPPSQTFPPPPSPSKSVAMQOKPTAYVVRPMGQQAQPS 307
 QY 58 PELDLIYGRKXKDD-----QTLDFYCIQPSVHVLRKSNWPEPCKPEPVCKVAMREF- 113
 DB 308 PEL-- - KPLPEYRQCTTEKTDKVPKAKUKLKMPSQSVEQYTSKEVHCVEIL 360
 QY 114 -RVJH- - -TALHSSSYREAVFMKLSKESLQIIVATFGLSSDPIALGVQDKDLF 166
 DB 361 KEMTHSWPEPTTAHTPTAEPKFP-PTKOS-----QHVS 396
 QY 167 SVFAPKMLLTLVPAHPALNAVLVHVSAGSAMPTGCTDSSSRMPSSSYRDMPGGLF 226
 DB 397 SVTONQYQYDTSSKTHS-----NSCQTSSMLEDDIQ----- 428
 QY 227 EGLSDDED-FHPNTRSTPSSSPASLGYSGAAGRPITOSELATALA--LASTPES 284
 DB 429 --LSDSEISDSEQTPEKPPSSSAPPSPQSL-----PEPVASAKSSSAESSESTSDSDS 479
 QY 285 SSHTPTRTTGHSSGTSPMSSGVQSCFTPIITDLF--SQALCHALCASQPSLQSQMQL 342
 DB 480 SSSSESSSSSDSENEPLETAPPEPPTTKWQDNLWTKVQPAAPPSPGPRSTPEPP 539
 QY 343 QQLRDMG:QDDELSPR-----PVTGTSKQWSSLSLEERH-----ELPSPPEP 389
 DB 540 RHPEKSGSDSATSQESESKEPPPSKKAFAFAP--EAPHGKRSQKSPAQOEPP 595
 RESULT 14
 PCT/US94-04496-26
 ? Sequence 26, Application PCT/US9404496
 ? GENERAL INFORMATION:
 ? APPLICANT: Crocco, Carlo
 ? APPLICANT: Canabadi, Eli
 ? TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
 ? TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
 ? TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
 ? NUMBER OF SEQUENCES: 86
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
 ? ADDRESS: No. 604214015
 ? STREET: One Liberty Place, 46th floor
 ? City: Philadelphia
 ? STATE: Pennsylvania
 ? COUNTRY: USA
 ? ZIP: 19103
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? OPERATING SYSTEM: IBM PC compatible
 ? SOFTWARE: Patent in Release #1.0, Version #1.25
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: PCT/US94/04496
 ? FILING DATE:
 ? CLASSIFICATION:
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: DeLuca Esq., Mark
 ? REGISTRATION NUMBER: 33,429
 ? REFERENCE/DOCKET NUMBER: TJU-1242
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (215) 568-3100
 ? TELEFAX: (215) 568-3439
 ? INFORMATION FOR SEQ ID NO: 26:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 1210 amino acids
 ? TYPE: amino acid
 ? TOPOLOGY: linear
 ? N-TERMINUS TYPE: Protein

PCT US94-04496-26

Query Match 5.7% Score 122.57 DB 5; Length 1210;
Best Local Similarity 21.1% Pred. No. 0.054;
Matches 88; Conservative 52; Mismatches 171; Indels 107; Gaps 18;

QY 8 LAKVLADQPTKSL-----RLEPTELEYSISLKLKOLIA-----GKLOESYVD 57
DB 249 LAKVKHDKLTPCDSLVAPAQPPGCTTPFTPSKSVAMCQKPTAYVRYGCGQOAPSES 307
QY 58 PELIDLIYGRKIKD---CTLDYGIOPGCVYVLRKSWPFDOKPKNKVAAYREF- 113
DB 308 PEL-----KLPDQYRQCTFERDLKVPALAKTKLKMPCSEQYQYEVHCVFEL 360
QY 114 -RVH-----TALHSSSSYREAVPKYLSNKESLQCTVAIFGLSEDLALGVLOOKLE 166
DB 361 KEMTWSRPPLTAIHTPTSAESYEP-PTTKUS-----QHVS 396
QY 167 SVFADPNMCLTLVPAHPALVNAIVLVHSAVSAAPMPGCTOSSSRMPSSSYECMPGGELP 226
DB 397 SVTONKQYDTSSKTHS-----NSQQTSSMLEDLQ----- 428
QY 227 EGLSDDED-DHFNTRSTSSSTSPSPASLSYGAAGPPPTQSELATALA-LASTPES 284
DB 429 ---LSSEDCSCSECTPKPSSSASAPSL-----PEPVASASSGASSESTSDUS 479
QY 285 SSGHTPTQGHSSSTSPMSSGVCTPTNLF SQALQHALCASCPESLQSQWQQL 342
DB 480 SSSSESSSSSSSENEPLETANUPERTTKWQDLNALTQVSGPAAPFEGPRSTEP 539
QY 343 QQLRMGICQDELSRPR-----FVGTQSKQWSSLSLEPH-----ELPASPEPP 382
DB 540 RHPEKSGSCSATSQSHSESKOFFKSSSKAPRATF-EAPHQKSKQKSPAQOQEP 595

RESULT 15

US-09-911-393-2

Sequence 2, Application US/529-11393

Patent No. 6323008

GENERAL INFORMATION:

APPLICANT: BELLETIER, Marc

APPLICANT: PARKER, William A.

APPLICANT: HAKES, David C.

APPLICANT: ZOFF, David A.

TITLE OF INVENTION: METHODS FOR PRODUCING

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & REMYUS LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/911,393

FILING DATE: 14-AUG-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7189-012-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)7909090

TELEFAX: (212)8699741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1060 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Protein

US-08-911-393-2

Query Match 5.7% Score 121; DB 4; Length 1060;

Best Local Similarity 22.1% Pred. No. 0.06;

Matches 101; Conservative 55; Mismatches 194; Indels 108; Gaps 21;

QY 6 WHLAVKIAD-----QPLTPKSLRLPE---TELGEYSLGGYSIS----- 41
DB 577 YHVVLTMANKIGSVYIDCEPLEGGQTVVPDGRTPDISHFYVGGYGRSDMPT-SHVTWN 636
QY 42 ---FLKCLIAKLCQESVPDELIDLIYCGRKLDQDQDLDFYGIOPGSTVHLRKSMPEPD 98
DB 637 VLLYRNLNAEIRTLFLSQDLIG-----TEAHMGSSSSSAH-----STPS-- 678
QY 99 OKPEPVKVAAYREFRVLHTALHSSSSYREAVFKLSNKESLDQIIVATP-----GLSS 152
DB 678 ---TPALNGAHS:TPSTPADSSAHS:TPS-----TPADSSAHS:TPSAPDNGAHS 723
QY 153 DPALGVCKDKJESVPADPNMLDTPVPAHPALVNAIVLVHVS-----AGSAPMPG 204
DB 724 TPSTPGRSSAHS:TPSTPADNGAHS:TPSAPADNSA-----HSTPSTPADNGAHS:TPSTP 776
QY 205 TDSRSSRYESS-----SYRDMFGGLFEGCL-SDDEDDFHFNTRSTPSSSTPSSRPSL 256
DB 777 ADNGAHS:TPSTPGDNGAHS:TPSTPGDSSAHS:TPSTPADNGAHS:TPSAPADNSAHS:TPSTP 836
QY 257 GYSGAAGPPPTQSELATALALASTP-ESSSH-TPT-PGTQG-HSSGTSPMSSGVQSGCTP 312
DB 837 GONGAHS-TPSAPAD-SNAHS:TPSTPADSSAHS:TPSAPDNGAHS:TPSAPADSSAHS:TPS 894
QY 313 ITNDLFSCALQHALQASGPESLQSQWQFQLOOLRDMGIQDELRLRCPRPVGTSKOPWS 372
DB 895 APGONGAHS:TPSAPADNGAHS:TPS-----APGDSNAHS:TPSTPADSSAHS:TPS 942
QY 373 SSLIEEHHELPAAP-EPPASCRGYCPWEALMKVPSPSLP 409
DB 943 TPADSSAHS:TPSAPDNGAHS:TPSAPADSSAHS:TPS:P 980

Search completed: November 12, 2003, 06:40:02

Job time : 44 secs